

CAPÍTULO 8.

HERRAMIENTAS DE SOFTWARE PARA OAI-PMH

Gema Bueno-de-la-Fuente (Univ. Carlos III)
David Rodríguez-Mateos (Univ. Carlos III)

Versión preprint de la obra publicada:

Bueno-de-la-Fuente, Gema; Rodríguez-Mateos, David. « Herramientas de software para OAI-PMH ». En: Bueno-de-la-Fuente, Gema (coord.) *La Iniciativa de Archivos Abiertos (OAI): situación y perspectivas en España y Latinoamérica*. Bogotá: Rojas Eberhard, 2007, pp. 247-302. ISBN 978-958-9121-89-4.

Sitio web del editor: http://www.rojaseberhard.com.co/rojaseberhard/bibliotecologia/lib_oai.html.

8.1. INTRODUCCIÓN

8.1.1. Herramientas software que soportan el protocolo OAI-PMH

Desde los orígenes de OAI-PMH en 1999, se han ido creando una serie de herramientas software que soportan distintos aspectos, roles y funciones de la arquitectura del protocolo, todas ellas con el objetivo final de facilitar la distribución de información científica y académica, así como de cualquier tipo de contenido electrónico, a través de la Web.

Para conocer estas herramientas resultan de gran utilidad los listados que ofrece tanto el sitio web de la propia Open Archives Initiative¹ como el proyecto europeo Open Archives Forum (OAF)². Los programas recogidos en estas fuentes se caracterizan por ser software libre y haber sido desarrollados por implementadores de la comunidad OAI, principalmente en el seno de proyectos de investigación en universidades pero también en centros de investigación, bibliotecas y consorcios bibliotecarios, proyectos de bibliotecas digitales e incluso por programadores individuales.

En el sitio web de la Open Archives Initiative se recogen unas 30 herramientas junto al enlace a su página web, el nombre de la organización que las ha desarrollado, una breve descripción de sus prestaciones y características, y la versión del protocolo OAI-PMH con el que cumplen. No obstante, este listado no resulta lo suficientemente completo, la información sobre cada herramienta viene a ser excesivamente escueta, puesto que no ha sido actualizado recientemente, se omiten algunas aplicaciones de relevancia³.

En cuanto al informe de OAF, ofrece información más detallada y acerca de un mayor número de herramientas (un total de 34), elaborando una ficha estándar en la que figuran: datos de control (como nombre de la herramienta, programador, organización, enlaces para su descripción y su descarga); información sobre su finalidad o propósito (grupo de usuarios, descripción, formatos, interfaces, sistema de colecciones, formatos de metadatos); requisitos técnicos fundamentales; soporte técnico que ofrecen (instrucciones para su instalación, datos de contacto, FAQ), y otros datos de su evaluación (reusabilidad, prestaciones adicionales, adaptabilidad, y aspectos de programación). Con “grupo de usuarios” diferencian entre proveedores de datos (DP – *data provider*) y proveedores de servicios (SP – *service provider*), aunque hay herramientas que cubren los dos aspectos del protocolo. El inconveniente principal de

¹ Listado de herramientas compatibles con el protocolo OAI-PMH, recogidas por la *Open Archives Initiative*: <<http://www.openarchives.org/tools/tools.html>>. [Fecha de consulta: 10/07/2007].

² Inventario de productos software para implementar archivos abiertos, de la *Open Archives Forum*: <<http://www.oaforum.org/resources/tvtools.php>>. [Fecha de consulta: 10/07/2007].

³ Algunos de estas herramientas han sido descritas en castellano por Teresa Silió (2005).

este informe se debe a que desde su publicación en 2003 se ha quedado desactualizado en algunas cuestiones, no contemplando nuevas versiones de herramientas ya existentes ni otros sistemas de aparición posterior a dicha fecha.

En estas fuentes podemos distinguir una tipología variada de herramientas que soportan el protocolo OAI-PMH, de las que se describe una amplia selección en los siguientes epígrafes. Estas aplicaciones van desde plataformas para la creación de repositorios desde su origen y que proporcionan una serie de funcionalidades básicas, como DSpace o EPrints (epígrafe 8.2), aplicaciones generadoras de proveedores de datos a partir de colecciones ya existentes (epígrafe 8.3), proveedores de servicios o recolectores metadatos en otros repositorios OAI (epígrafe 8.4), y otras herramientas muy diversas que soportan el protocolo OAI-PMH en sus múltiples roles (epígrafe 8.5).

Las herramientas que se describen en este capítulo son en su totalidad software libre, ya que, aunque existe un considerable número de aplicaciones comerciales para la implementación de archivos abiertos, la mayor parte son adaptaciones o servicios añadidos a otros sistemas de software libre así como meros módulos a sus diversos productos y sistemas de gestión de información.

8.1.2. Software libre y software propietario

8.1.2.1. Los movimientos de software libre y software de código abierto

Aunque en la literatura científica a menudo se utilizan de forma indistinta las expresiones “software libre” o *free software* y “software de código abierto” (software abierto) u *open source software*, es posible establecer ciertos matices entre estos conceptos, que han surgido en el seno de movimientos diferentes aunque paralelos.

La denominación “software libre” proviene del movimiento Free Software Foundation (FSF) iniciado en los años 80 y que se refiere al software que garantiza las libertades de utilización, ejecución, copia, estudio, redistribución y mejora del programa. El concepto *free*, que en inglés significa libertad y gratuidad, generaba ciertas confusiones, por lo que algunos buscaron un término que las solventase.

A principios de los 90 surgió el movimiento por el *open software* de código abierto liderado por el Open Software Initiative (OSI), que pone el acento sobre el derecho al acceso al código fuente del programa, estando permitida su redistribución, modificación y derivación de cara a una mayor calidad de las aplicaciones. El apelativo de “abierto” es criticado por poder ser interpretado como un término más débil y flexible que el de “libre”.

En la actualidad ambos movimientos, aunque incompatibles desde el punto de vista filosófico, son completamente equivalentes desde el punto de vista práctico; de

hecho, ambos movimientos trabajan juntos en el desarrollo práctico de proyectos⁴. Por ello, emplearemos el término “software libre” para referirnos a todos los sistemas no propietarios descritos a lo largo de este capítulo.

8.1.2.2. *Software libre frente a software comercial*

La mayor parte de los sistemas disponibles para la implementación de archivos abiertos son software libre, y en general, son los más recomendados por los expertos. Además de ser gratuitos en su adquisición, generalmente dependen de otras tecnologías también libres (tecnología Java, servidores Apache, bases de datos MySQL, etc.), y se basan en estándares internacionalmente aceptados en cuanto a formatos de archivos y esquemas de metadatos. Lo más destacado del software libre es que se beneficia de la aportación de múltiples programadores de todo el mundo, que mejoran el sistema continuamente o lo amplían con nuevas funcionalidades.

No obstante, no se deben pasar por alto algunos inconvenientes que se pueden presentar al utilizar uno de estos sistemas. Cuando se elige un software libre, la institución debe ser capaz de instalar, adaptar, gestionar y mantener en funcionamiento el sistema por sí misma, ya que, al contrario que ocurre con los productos comerciales, no se ofrece un servicio técnico y de apoyo a los usuarios del software por parte de los desarrolladores. Hay que asegurarse de que se cuenta con personal capacitado para llevar a cabo estas tareas, y por supuesto, que esté familiarizado con sistemas operativos como Linux, también de código abierto, ya que normalmente las aplicaciones *open source* se diseñan para funcionar en estos entornos, y en menor medida para sistemas operativos comerciales como Microsoft Windows. En ocasiones, el gasto en contratación o formación de personal puede superar con creces el de adquisición de un producto comercial.

En el caso del software propietario, en general las empresas proporcionan un software genérico, que en muchas ocasiones se basa en algún otro software libre ya existente, y posteriormente lo implementan y adaptan de forma específica en diferentes instituciones dependiendo de las necesidades o requerimientos de éstas. Es el caso de productos como Open Repository de PubMed, basado en el software libre DSpace⁵, o la versión comercial de la plataforma Fedora, distribuida bajo la marca VITAL por la compañía VTLS (Vision Technology in Library Solutions) (Veáse el epígrafe 8.2.3. *Fedora*).

⁴ Resultan muy ilustrativos los artículos publicados en Wikipedia sobre software de código abierto <http://es.wikipedia.org/wiki/Codigo_abierto> y software libre <http://es.wikipedia.org/wiki/Software_libre>. [Fecha de consulta: 10/07/2007].

⁵ Open Repository: <<http://www.openrepository.com/>> es un servicio de BioMed Central para construir, lanzar y mantener repositorios institucionales para organizaciones, basado en el software DSpace. Principalmente diseñado para universidades e instituciones de investigación, está implantado en países como Francia, Italia, Reino Unido, Estados Unidos, e Islandia. [Fecha de consulta: 10/07/2007].

La estrategia fundamental de estos productos comerciales se basa en la creación de contratos de servicio que aseguran a las instituciones no sólo la instalación del sistema, sino también diversos servicios como el mantenimiento, actualización, mejora y realización de revisiones de forma periódica, la asistencia en caso de fallos técnicos, la generación de copias de seguridad e incluso el mantenimiento de los servidores en los que se alojan los repositorios implementados.

Otros servicios de valor añadido se basan en la configuración y adaptación de los interfaces del repositorio a la imagen institucional del cliente, la creación de logotipos personalizados, el soporte a una mayor variedad de formatos de archivo y de esquemas de metadatos, o el envío de estadísticas de uso y descarga de documentos. En resumen, se persigue una mayor sencillez de utilización de cara a los usuarios, una despreocupación en los aspectos técnicos para los administradores; y una mayor y más rápida percepción de éxito del sistema por parte de los gestores.

Por último, cabe destacar diversos productos de software comercial para la creación de bibliotecas digitales, que no son meras adaptaciones de otros sistemas de software libre, y que ya soportan el protocolo OAI-PMH (algunos mediante plug-ins o módulos adicionales), permitiendo el establecimiento de proveedores de datos. Algunas de las herramientas más conocidas son: CONTENTdm⁶, Digital Commons⁷ (antes bePress), Digitool de Exlibris⁸, Curator de Endeavor⁹, Archivemaker de digitAlexandria¹⁰, DLXS¹¹ de la Universidad de Michigan, así como la plataforma Intralibrary de Intrallect¹², que aunque está más bien enfocada a los recursos educativos u “objetos de aprendizaje”, puede ser empleada para establecer un repositorio institucional.

⁶ CONTENTdm es un paquete de software comercial de DiMeMa <<http://www.dimema.com>> en colaboración con el consorcio bibliotecario OCLC, que proporciona herramientas para administrar, organizar, publicar y realizar búsquedas en colecciones digitales, y que entre sus funcionalidades incluye la compatibilidad con el protocolo OAI-PMH. [Fecha de consulta: 10/07/2007].

⁷ Digital Commons es un software de Publicaciones Electrónicas Berkeley (BePRESS) <<http://www.bepress.com/digitalcommons.html>> para crear repositorios institucionales compatibles con el protocolo OAI-PMH 2.0. [Fecha de consulta: 10/07/2007].

⁸ Digitool de Exlibris es una solución empresarial para la gestión de activos digitales en bibliotecas y entornos académicos: <<http://www.exlibrisgroup.com/digitool.htm>>. [Fecha de consulta: 10/07/2007].

⁹ Esta información se puede extraer del folleto del producto Curator de Endeavor, para la gestión de activos digitales en bibliotecas: <<http://www.endinfosys.com/software/brochures/curator.pdf>>. [Fecha de consulta: 10/07/2007].

¹⁰ ArchiveMaker <<http://www.bdaweb.net/template/inc/archivemaker/>> es una herramienta para crear repositorios OAI de forma sencilla, creada por digitAlexandria. [Fecha de consulta: 10/07/2007].

¹¹ Digital Library Extensión Service (DLXS) <<http://www.dlxs.org/products/products.html>>, un producto para la construcción de bibliotecas digitales diseñado por la Universidad de Michigan, y utilizado como base del proveedor de servicios OAIster. Soporta el protocolo OAI-PMH mediante un módulo CGI llamado broker20. [Fecha de consulta: 10/07/2007].

¹² Intralibrary es el software de la empresa Intrallect <<http://www.intrallect.com/>> para construir un repositorio de objetos de aprendizaje para cualquier institución o proyecto. Permite crear cuatro tipos de aplicaciones: repositorio de objetos de aprendizaje SCORM, sistema de gestión del contenido educativo (LCMS), sistema de biblioteca digital, repositorio nacional o regional, etc. [Fecha de consulta: 10/07/2007].

Un caso especial es el del sistema polaco dLibra (<http://dlibra.psnc.pl/>)¹³, creado por el Poznan Supercomputing and Networking Center (PSNC), de uso generalizado en Polonia para la construcción de bibliotecas digitales conformes al protocolo OAI.

Por último, algunos sistemas comerciales para la gestión de activos digitales, como ArchivalWare (<http://www.ptfs.com/>), o para la gestión de archivos de empresa, como ADLIB Archive (<http://www.adlibsoft.com/>), también soportan el protocolo OAI-PMH.

Todos estos desarrollos demuestran la fuerte relación existente entre bibliotecas digitales, gestión de contenidos y repositorios, más aún en el marco del movimiento de archivos abiertos y en el uso de protocolos y estándares de interoperabilidad que fomentan el intercambio de información entre sistemas.

¹³ dLibra Digital Library Framework: <<http://dlibra.psnc.pl/>>, un desarrollo del Poznan Supercomputing and Networking Center (PSNC): <<http://www.man.poznan.pl/>>. [Fecha de consulta: 10/07/2007].

8.2. PLATAFORMAS PARA LA CREACIÓN DE REPOSITORIOS OAI

Se entiende por plataformas para la creación de repositorios OAI aquellos paquetes de software dirigidos a la implementación integral de archivos o repositorios que cumplan con el protocolo OAI-PMH. La mayor parte son herramientas *out-of-the-box*, o paquetes distribuibles que no necesitan excesiva configuración o adaptación por parte de los administradores, aunque en ocasiones precisen de programas complementarios, por ejemplo, para la implementación de interfaces¹⁴.

Estos sistemas, una vez instalados y configurados en un servidor web determinado, deben ofrecer una serie de funcionalidades básicas para el mantenimiento y gestión del repositorio. Una de sus funcionalidades principales de gestión es el soporte al flujo documental (remisión de documentos; evaluación, aceptación o rechazo de documentos); la edición, revisión y evaluación de metadatos sobre los recursos; y la transformación de formatos de fichero, entre otras. En general, ofrecen interfaces de usuario, de autor y de administrador, y permiten la creación de diversos grupos de usuario y niveles de acceso. En cuanto al contenido, facilitan la creación de colecciones de materiales y generalmente aceptan múltiples formatos de archivo. Las funcionalidades relativas al uso del archivo se limitan a diversas opciones de búsqueda y navegación entre los registros del repositorio (palabras clave, búsqueda por campos de metadatos), así como funciones para su visualización y descarga.

En relación a estas herramientas o sistemas software que hemos dado en llamar "plataformas para la creación de repositorios OAI", es posible señalar algunos recursos de interés que ofrecen información sobre las mismas, y que han servido de punto de partida para el análisis propuesto.

Uno de los recursos más citados al hablar de estas herramientas es *A Guide to Institutional Repository Software* (Crow, 2004), una guía de software para repositorios institucionales, elaborada por el Open Society Institute para ayudar, a aquellas organizaciones que quieran utilizar algunas de las herramientas disponibles, a seleccionar la que más se ajuste a las necesidades de su proyecto de repositorio. En la tercera versión de la guía, publicada en agosto del 2004, se analizan un total de nueve sistemas que satisfacen tres criterios básicos: cumplir la última versión del protocolo OAI-PMH (v2.0), ser de código abierto, y estar disponibles de forma pública y gratuita para su descarga e implementación. Estos nueve programas son *Archimede*, *ARNO*, *CDSware*, *DSpace*, *Fedora*, *Eprints*, *i-Tor*, *Mycore* y *OPUS*.

La guía se divide en dos partes: en la primera se presentan brevemente las características de cada sistema, incluyendo datos de contacto y enlaces para información adicional, y en la segunda se presenta una tabla en la que se analizan de forma detallada las funciones y características de los programas. Hay que tener presente

¹⁴ Este es el caso de herramientas como Fedora y Arno, que se analizarán más adelante, y que no ofrecen un interfaz predefinido y es necesario utilizar otros sistemas complementarios y establecer una configuración o arquitectura propia.

que cada sistema ha sido desarrollado con distintas filosofías y para cumplir objetivos diferentes, por ello, sus características variarán en gran medida de uno a otro haciendo que sean más o menos adecuados para los proyectos de otras instituciones.

Otros recursos de gran utilidad para la creación de un repositorio institucional y que analizan distintos sistemas software, son el cuaderno de trabajo LEADIRS¹⁵ elaborado por la Biblioteca del Massachussets Institute of Technology (MIT), o informes como el publicado por Gibbons en la revista *Library Technology Report*¹⁶. El primero analiza un total de nueve sistemas, seis de ellos software libre (Archimede, CDSWare, DSpace, EPrints, Fedora, Greenstone) y tres comerciales (bePress, CONTENTdm y Open Repository). Por su parte, el informe de Gibbons recoge once sistemas, nueve en la categoría de software libre (Arno, CDSware, DSpace, EPrints, Fedora e i-Tor, como sistemas ya consolidados, y MyCore, eDoc-Server y OPUS, como sistemas en desarrollo), junto a los productos comerciales bePress y Documentum. También menciona brevemente sistemas específicos según su función, como el VT ETD-db (Virginia Tech), dirigido a repositorios institucionales de tesis y disertaciones electrónicas; así como ePortfolios de la Universidad de Minnesota, Duluth, y otros proyectos de la OSPI (*Open Source Portfolio Initiative*), en cuanto a sistemas que soportan la creación de portafolios electrónicos.

	SOFTWARE LIBRE	SOFTWARE COMERCIAL
SOFTWARE GENÉRICO	Archimede ARNO CDSware DSpace eDoc-Server EPrints Fedora i-Tor MyCoRe OPUS	CONTENTdm Documentum Open Repository Digital Commons Curator DigiTool
SOFTWARE ESPECÍFICO	VT ETD-db DoKs ePortfolio	bePress IntralllectIntralibrarv

Figura 8.1. Algunas plataformas de software para repositorios

¹⁵ Mary R. Barton y Margaret M. Waters. *Creating an institutional repository: LEADIRS workbook*. Cambridge, MA: MIT Library, 2004. Disponible en: <<http://www.dspace.org/implement/leadirs.pdf>> [Fecha de consulta: 10/07/2007]

¹⁶ Susan Gibbons. "Establishing an Institutional Repository". *Library Technology Report*, vol. 40, n. 4, 2004, p. 11-14.

Los sistemas mencionados se podrían clasificar según sean software libre o software comercial por un lado, y por ser sistemas específicos para una función (por ejemplo, para una tipología documental concreta como las tesis electrónicas) o sistemas genéricos con diversas funciones y aplicables a todo tipo de materiales y finalidades.

De los sistemas software conocidos, se han seleccionado, en primer lugar, aquellos que cumplen las tres características básicas que ya determinó la guía elaborada por el Open Society Institute: cumplir la última versión del protocolo OAI-PMH, ser de código abierto, y estar disponibles de forma pública y gratuita para su descarga e implementación; estos son (exceptuando i-Tor¹⁷): los ocho siguientes: Archimede, ARNO, CDSware, DSpace, Fedora, EPrints, Mycore y OPUS. Todos son plataformas para la creación de repositorios abiertos de carácter genérico, es decir, no definidas por su tipo de documentación, finalidad o formatos de archivo. Junto a éstas, se ha incluido en el análisis algunas herramientas consideradas software específico como ETD-db o DoKs, ambos dirigidos a la gestión y publicación de tesis electrónicas, así como el sistema de construcción de archivos abiertos personales Kepler.

Se han analizado otros sistemas OAI no recogidos en la literatura científica o en listados de software, sino que se han localizado a partir de los registros de proveedores de datos ROAR y OpenDOAR. Un caso particular ha sido el de Wildfire (<http://wildfire.eldoc.ub.rug.nl/>), una plataforma software para construir repositorios abiertos creada por Henk Druiven de la Universidad de Gröningen, Holanda. Funciona sobre PHP y MySQL en un servidor Apache. Una de sus peculiaridades es que permite la construcción de agregadores de proveedores de datos, y la universidad ha establecido varios de ellos. En su sitio web se indica que se emplea en 70 repositorios, y en OpenDoar figuran 23 de ellos, lo que supone más de la mitad de los repositorios holandeses recogidos por este directorio. No obstante, todos ellos son repositorios de facultades, escuelas, institutos y otros centros de la propia universidad de Gröningen, y no figura ningún repositorio externo a dicha institución. Éste sistema no ha sido incluido en el estudio porque no está disponible para su libre descarga ni se ofrece el código fuente del programa, y por lo tanto no puede considerarse software libre.

Muy similar es el caso del sistema de publicación electrónica de las universidades de Suecia y Noruega (DiVA). Aunque es un producto distribuido de forma interna entre las instituciones que participan en el proyecto, y no se puede considerar como software libre (ni tampoco es un producto comercial), se determinó necesaria su inclusión, por ser el sistema de repositorio institucional más representativo de los países escandinavos, con un éxito comparable a Opus en Alemania y Arno en Holanda.

¹⁷ En cuanto a la herramienta i-Tor (Tools and Technologies for Open Repositories) <<http://www.i-tor.org>>, sistema de repositorio de código abierto iniciativa del Instituto Holandés de Servicios de Información Científica (NIWI), se ha desestimado su descripción ya que, además de contar con escasas implementaciones que en su mayor parte son servicios web distintos de los repositorios, ya no se encuentra activa la página web del proyecto. [Fecha de consulta: 10/07/2007]

El total de doce herramientas analizadas se presentan a tenor de un laxo criterio de difusión geográfica y/o numérica¹⁸. En primer lugar, se describen los sistemas de carácter general con mayor difusión a nivel internacional, como son GNU EPrints y DSpace (ambos con más de 200 implementaciones), seguidos de Fedora, que aunque sí tiene difusión mundial todavía se le conocen muy pocas implementaciones. A estos tres se añade CDS Invenio, de alcance europeo.

En cuanto a los sistemas de difusión nacional, destacan OPUS y MyCoRe, que se emplean únicamente en universidades de Alemania; ARNO, el sistema por excelencia de los repositorios de universidades holandesas; DiVa, adoptado por distintas universidades de los países escandinavos, Noruega y Suecia; y por último Doks, dirigido a las instituciones de enseñanza superior de la región flamenca de Bélgica. En cuanto a Archimède, tiene una limitada difusión institucional ya que de momento sólo se ha implementado en la Universidad de Laval en Canadá.

Para finalizar, se describen los sistemas de carácter específico Kepler y ETD-db (DoKS, como ya se ha comentado, es de aplicación belga), ambos con una cierta difusión internacional, si bien el último es mayoritariamente utilizado por instituciones estadounidenses.

8.2.1. GNU EPrints

GNU EPrints (antes EPrints) (<http://www.eprints.org>) ha sido desarrollado en la Universidad de Southampton y actualmente está financiado por JISC (Joint Information Systems Committee) del Reino Unido, como parte del Open Citation Project, y por la NSF (National Science Foundation) de EEUU. En principio está diseñado para la creación de repositorios de e-prints, aunque puede ser empleado con otros objetivos. Hasta el momento, el software GNU EPrints es el segundo sistema más utilizado en la creación de repositorios (DSpace le ha superado recientemente), con más de doscientas implementaciones que suman más de trescientos mil registros. El gran número de instalaciones de GNU EPrints sugiere que una institución puede instalarlo y ponerlo en marcha de forma relativamente rápida, y con un mínimo de habilidad técnica.

¹⁸ Un trabajo previo de Gema Bueno de la Fuente, en el doctorado de Documentación de la Universidad Carlos III de Madrid, analiza la difusión de las herramientas software para la creación de archivos abiertos recogidas en guía elaborada por el Open Society Institute, en base a criterios como número de implementaciones, países, número de registros, tipos de repositorio y otras cuestiones. La toma de datos se llevó a cabo en junio de 2005, y se partió de los datos ofrecidos por el registro de archivos abiertos ROAR <<http://roar.eprints.org>>, estos fueron complementados con la inclusión de otros repositorios no contemplados en el registro y con consulta directa de todos los archivos comprendidos en el estudio, un total de 316. Las conclusiones que se obtuvieron de la realización de dicho estudio han permitido realizar las afirmaciones que se recogen en este epígrafe.

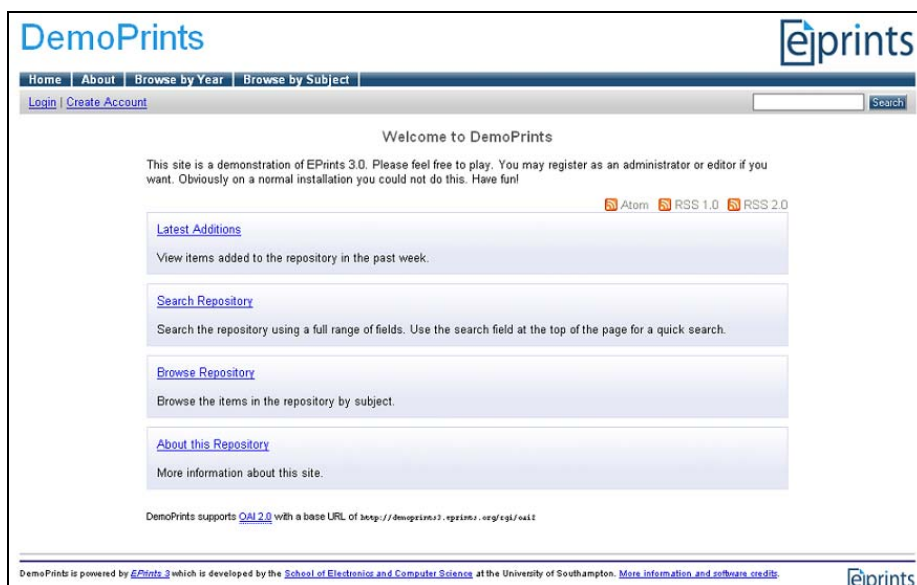


Figura 8.2. Página principal de la Demo de GNU Eprints 3 [<http://demoprints3.eprints.org>]

Actualmente en su versión 3.0, el código de E-prints está escrito en Perl, utiliza el servidor web Apache y MySQL como base de datos, pudiendo funcionar sobre los sistemas operativos Linux, GNU o Solaris, así como en Windows gracias a un módulo adicional desarrollado su última versión.

Aunque por defecto se emplean metadatos Dublin Core, EPrints permite aplicar cualquier esquema de metadatos, y que el administrador decida qué campos se asignan a cada tipología documental. EPrints permite la recuperación de contenido mediante búsqueda textual (simple o avanzada, variando el número de campos de metadatos en los que se puede consultar) y navegación.

En cuanto a formato, un archivo EPrints puede almacenar cualquier tipo de ficheros que el administrador decida aceptar, incluyendo HTML, Adobe PDF, y PostScript, y cada documento puede ser almacenado en varios formatos distintos. El proceso de remisión de contenido contempla las tareas de evaluación y revisión de metadatos, comprobación y conversión de formatos, y las opciones de aprobación, rechazo o reenvío del documento al autor para su modificación. GNU Eprints permite crear identificadores persistentes, aunque se recomienda asignar identificadores propios independientemente del sistema.

En general, se considera que este sistema funciona mejor para documentación textual y no multimedia, y no soporta formatos para objetos complejos como METS, MPEG21 o IMS CP (para el empaquetamiento de objetos de aprendizaje), por lo que no se suele emplear por bibliotecas digitales, repositorios de objetos de aprendizaje, ni por aquellos repositorios que pretendan albergar documentación gráfica, sonora o audiovisual.

8.2.2. DSpace

DSpace (<http://www.dspace.org>) ha sido desarrollado conjuntamente por los laboratorios Hewlett-Packard y por la biblioteca del MIT (Massachusetts Institute of Technology) específicamente para constituir servicios de repositorio institucional. DSpace es un sistema diseñado para la captura, almacenamiento, indización, preservación y redistribución de la producción intelectual del personal investigador de la universidad en formato digital. Disponible desde noviembre del 2002, y distribuido mediante una licencia BSD (*Berkeley Software Distribution*), en mayo del 2007 ha publicado su versión 1.4.2¹⁹.



Figura 8.3. Página principal del repositorio e-Archivo de la Universidad Carlos III de Madrid, construido con DSpace [<http://e-archivo.uc3m.es:8080/dspace>]

DSpace está construido en base al concepto de comunidades y subcomunidades de una institución y sus colecciones. Las comunidades representan unidades organizativas, como departamentos académicos o administrativos, centros de investigación, laboratorios e incluso investigadores individuales. Además DSpace organiza los usuarios en cuentas personales (*e-person*)²⁰ y grupos, obligando a todos los usuarios personales a pertenecer a un grupo definido para poder asignarle permisos de

¹⁹ La versión más reciente de DSpace puede encontrarse en: http://sourceforge.net/forum/forum.php?forum_id=594714. [Fecha de consulta: 10/07/2007]

²⁰ Dspace denomina a los usuarios e-persons porque algunos usuarios pueden ser máquinas en vez de personas.

escritura, lectura, modificación y eliminación de registros. Estas autorizaciones pueden realizarse a nivel de comunidad o colección.

DSpace está programado en Java y corre sobre sistemas tipo Unix, aunque ya se han realizado con éxito pruebas de instalación en Windows XP. Precisa de un servidor web Apache y una base de datos relacional de código abierto PostgreSQL 8.0 o superior, permitiendo también el uso de Oracle 9.0 o posterior²¹.

Aunque soporta todo tipo de formatos de documentos, recomienda a los administradores de cada sistema decidir qué formatos aceptar con fines de preservación, pudiendo diferenciar así tres categorías de formatos: soportados, conocidos y no soportados. Otra característica técnica de DSpace es el empleo de identificadores persistentes de recursos mediante el sistema Handle²² del CNRI (Corporation for National Research Initiatives). Además, el interfaz web básico que ofrece DSpace por defecto es totalmente configurable según las necesidades de cada institución o comunidad de usuarios, mediante HTML y Java.

El esquema de metadatos utilizado para la descripción es Dublin Core cualificado, basado en el perfil de aplicación de Dublin Core para bibliotecas²³, que puede ser adaptado a través del interfaz de administración. DSpace utiliza el motor de búsqueda de software libre Lucene, permitiendo búsquedas tanto por campos de metadatos como en el texto completo, así como realizar búsquedas booleanas.

8.2.3. Fedora

Fedora (*Flexible Extensible Digital Object and Repository Architecture*) (<http://www.fedora.info>) ha sido elaborado conjuntamente por la Universidad de Virginia y la Universidad de Cornell (EEUU) con un proyecto que comenzó en 1998 con financiación de la Fundación Andrew W. Mellon. El proyecto Fedora pretende proporcionar un software libre para repositorios digitales que puedan emplearse como fundamento para distintos tipos de sistemas de gestión de información: repositorios institucionales, bibliotecas digitales y otros sistemas de gestión de contenido, gestión de activos digitales, publicación académica y preservación digital.

Aún en desarrollo (su última versión es la 2.2²⁴), su nota más característica es que facilita cuatro interfaces o servicios web API (*API-A*, *API-A-LITE*, *API-M*, *API-M-LITE*), junto a otros tres para la búsqueda y exposición de metadatos con OAI-PMH (*Search*, *Simple OAI*, *RISearch*). En cuanto a los interfaces API, consisten en: *API-A*: Fedora

²¹ Para más información al respecto consulte la documentación técnica de DSpace: <<http://dspace.org/technology/system-docs>>. [Fecha de consulta: 10/07/2007]

²² Véase: <<http://www.handle.net/>> para más información sobre el sistema CNRI Handle. [Fecha de consulta: 10/07/2007]

²³ DCMI-Libraries Working Group. *DC-Library Application Profile (DC-Lib)*. Septiembre 2004. Disponible en: <<http://www.dublincore.org/documents/library-application-profile>>. [Fecha de consulta: 10/07/2007].

²⁴ Las funcionalidades de la nueva versión del sistema software Fedora se recogen en la página: <<http://www.fedora.info/download/2.2/userdocs/distribution/release-notes.html>>, destacando la compatibilidad con Java 1.5, su mayor facilidad de instalación y su empaquetamiento de distribución que permite que sea instalado como cualquier aplicación web en un contenedor de servlet. [Fecha de consulta: 10/07/2007].

Access Service (SOAP) que facilita el descubrimiento y disseminación de objetos en el repositorio; *API-A-LITE*: Fedora Access Service (REST), versión ligera del sistema de acceso implementada como un servicio disponible vía HTTP con REST²⁵; *API-M*: Fedora Management Service (SOAP) que define un interfaz para administrar el repositorio, incluyendo las operaciones necesarias para que los clientes puedan crear y mantener objetos digitales; y *API-M-LITE*, versión ligera del interfaz API-M con REST.

Fedora proporciona importantes funciones para la gestión de los repositorios, en cuanto al almacenamiento de ficheros, importación y exportación en XML y otros formatos de codificación como Fedora Object XML (FOXML)²⁶, e incluso una versión del esquema Metadata Encoding and Transmission Standard (METS)²⁷; control de versiones y establecimiento de relaciones entre objetos mediante el empleo de metadatos; control y autenticación de acceso mediante XACML²⁸; identificadores persistentes tipo DOI y URI's, así como utilidades para la migración, procesamiento por lotes y creación de informes.

Para la búsqueda y localización de recursos en sus repositorios, Fedora crea automáticamente dos índices: el índice de búsqueda por defecto permite buscar por las características de los objetos y elementos Dublin Core; y desde Fedora 2.0 se ofrece un nuevo índice de recursos basado en RDF que incluye más información sobre objetos y relaciones entre objetos. Ambos índices son buscables mediante los interfaces de servicio web tipo REST.

Fedora está programado en Java, y permite emplear Apache, MySQL y otro software de código abierto. Funciona sobre diversos sistemas operativos como Windows 2000, NT y XP, Solaris y Linux. VTLS (Vision Technology in Library Solutions) ofrece una versión comercial de Fedora, denominada VITAL²⁹, con extensiones de diverso tipo así como soporte técnico a las implementaciones del mismo (tanto de VITAL como de Fedora).

²⁵ REST es un término acuñado por Roy Fielding en su tesis doctoral, para describir un estilo de arquitectura de sistemas en red. REST es el acrónimo de Representational State Transfer. Fielding, Roy Thomas. *Architectural Styles and the Design of Network-based Software Architectures*. [En línea] Universidad de California, Irvine, 2000. 180 p. Tesis doctoral. Disponible en: <http://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf>. [Fecha de consulta: 10/07/2007]

²⁶ La definición del esquema Fedora Object XML (FOXML), está disponible en: <<http://www.fedora.info/definitions/1/0/foxml1-0.xsd>>. [Fecha de consulta: 10/07/2007].

²⁷ La definición de la extensión Fedora al esquema Metadata Encoding and Transmission Standard (METS), es: <<http://www.fedora.info/definitions/1/0/mets-fedora-ext.xsd>>. [Fecha de consulta: 10/07/2007].

²⁸ Incluye un módulo de control y autenticación de acceso que refuerza las políticas de control de acceso de "grano fino" empleando XACML (*eXtensible Access Control Markup Language*), un nuevo lenguaje basado en XML para la protección de datos empresariales. XACML ha sido desarrollado por SUN y aceptado por el consorcio OASIS, disponible en: <<http://sunxacml.sourceforge.net/>>. [Fecha de consulta: 10/07/2007].

²⁹ Para más información sobre el producto VITAL de VTLS (Visionary Technology for Library Solutions), véase: <<http://www.vtls.com/Products/vital.shtml>>. [Fecha de consulta: 10/07/2007].

8.2.4. CDS Invenio

CDS Invenio (<http://cdsware.cern.ch/invenio/>), llamado CDSWare hasta julio de 2006, es un desarrollo de la Organización Europea para la Investigación Nuclear o CERN (*Organisation Européenne pour la Recherche Nucléaire*) con sede en Ginebra, Suiza. Ha sido diseñado para actuar de forma simultánea como un servidor electrónico de preprints, un catálogo de biblioteca en línea con registros bibliográficos, e incluso como un sistema documental web. En el CERN Document Server (CDS), CDS Invenio gestiona unas 500 colecciones de datos, con más de 900.000 registros, incluyendo más de 360.000 documentos a texto completo (preprints, artículos, libros, revistas, fotografías y otros).

CDS Invenio puede correr sobre Linux o Unix, y depende de Apache, MySQL, PHP, Python y WML (Website META Language). Asimismo, el interfaz web es fácilmente configurable y ofrece opciones de personalización a los usuarios. Distribuido mediante una licencia GNU, en febrero de 2007 ha publicado su versión v0.92.1, y aunque el código de CDS Invenio es libre, CERN ofrece la opción de realizar un contrato de instalación y mantenimiento con carácter anual.

Al contrario que otros sistemas de archivos abiertos que utilizan generalmente Dublin Core sin cualificar, CDS Invenio emplea el esquema de metadatos MARC21. El sistema ofrece otras funcionalidades de conversión de documentos en diferentes formatos como HTML o XML en el momento del depósito, almacenando siempre el formato original de los documentos junto a diversos formatos de conversión indicados.

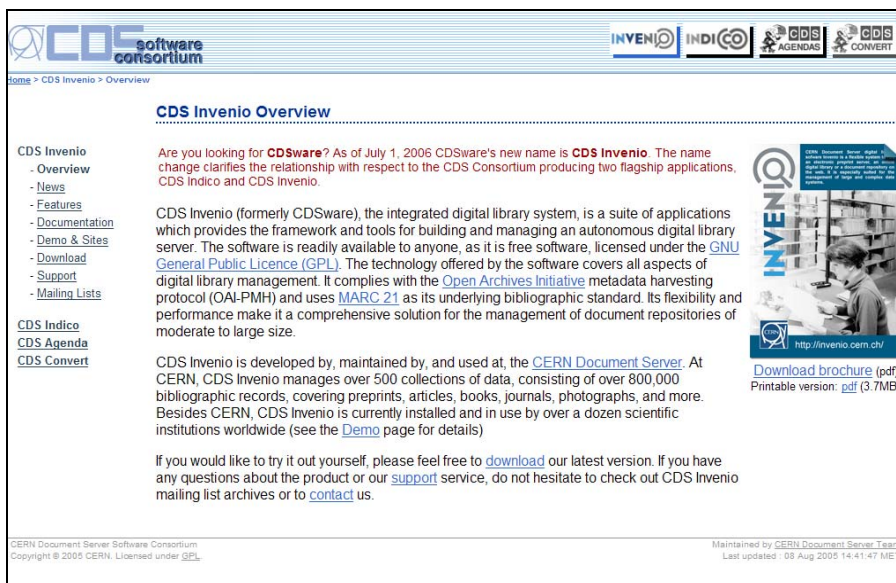


Figura 8.4. Página principal de CDS Invenio [<http://cdsware.cern.ch/invenio/>]

CDS Invenio cuenta con un potente motor de búsqueda propio, cuyo interfaz de búsqueda ha sido diseñado según el modelo de Google. Todos los campos de metadatos MARC son recuperables y navegables, y se permite la búsqueda a texto completo en formatos de ficheros PostScript, PDF, Microsoft Word, Excel y PowerPoint. Además, todas las referencias bibliográficas de los documentos son indizadas de forma separada, de manera que no sólo se puede buscar en las citas de referencia, sino que además el sistema permite identificar cualquier documento que cite el documento original.

8.2.5. OPUS

OPUS (Online Publikationsverbund der Universität Stuttgart) (<http://elib.uni-stuttgart.de/opus>) fue creado en 1998 como resultado de un proyecto de la biblioteca y el centro computacional de la Universidad de Stuttgart, Alemania. En la actualidad, esta universidad continúa su desarrollo con la colaboración del centro de servicio de bibliotecas de Baden-Württemberg (BSZ), y en febrero de 2006 lanzó su versión 3.0.



Figura 8.5. OPUS en la Universidad de Stuttgart (<http://elib.uni-stuttgart.de/opus/>)

OPUS es un software muy extendido en el ámbito universitario alemán, ya que lo utilizan más de 60 universidades alemanas para gestionar sus publicaciones electrónicas, incluyendo tesis y disertaciones y otros documentos académicos. La Universidad de Stuttgart ofrece en la página web de OPUS un interfaz de búsqueda simultánea en todos o algunos (a elección) de los repositorios que emplean este sistema³⁰.

Programado en PHP, este software corre sobre sistemas Linux o Solaris, y cualquier tipo de servidor web. OPUS admite cualquier formato de ficheros y ofrece la función del administrador para seleccionar los formatos aceptados. Permite convertir documentos electrónicos a PDF asignando automáticamente algunos metadatos que se almacenan en formato Dublin Core sin cualificar empleando una base de datos MySQL. La recuperación tiene por tanto las funciones típicas de otras bases de datos en línea: varios campos de búsqueda, opciones de truncamiento y operadores booleanos.

OPUS permite realizar importación y exportación de ficheros y metadatos, su interfaz es fácilmente configurable a la imagen de la institución que lo implementa, y permite obtener fácilmente estadísticas e informes de uso del sistema. En la versión 3.0 de OPUS se observa una separación del contenido y de las secciones del programa con todos los aspectos externos (independencia del idioma), al tiempo que se incluyen funcionalidades para la definición de restricciones o niveles de acceso a los documentos (mundo/campus/grupo predefinido), así como para la importación de datos bibliográficos sin incluir los textos completos de los documentos.

OPUS 3.0 posee un interfaz OAI que permite pasar de formatos de datos DC simple a formatos XMETADISS+, XMETADISS y XEPICUR. XMETADISS+ se utiliza para el intercambio de los datos con el grupo de bases de datos del centro de servicio de la biblioteca de Baden-Rurtemberg, empleando WinIBW para la importación. Las principales funciones para futuros desarrollos se centran en servicios de repositorios interoperables, como estadísticas de uso y análisis de citaciones en una red de repositorios certificados.

8.2.6. MyCoRe

El sistema MyCoRe (<http://www.mycore.de>) se desarrolla en el marco del proyecto MILESS de la Universidad de Essen, Alemania. MyCoRe es resultado de la colaboración de un consorcio de universidades para ofrecer un conjunto de herramientas software que permitan crear bibliotecas digitales y soluciones archivísticas (Content Repositories-CoRe) que sean configurables y adaptables a los requisitos y necesidades locales de cada sistema (My, para su adaptación local) sin esfuerzos de programación. Ya está disponible la versión 1.3.0 de MyCoRe desde junio del 2006.

³⁰ Búsqueda simultánea en todos los repositorios que emplean el software OPUS, en: http://elib.uni-stuttgart.de/opus/gemeinsame_suche.php. [Fecha de consulta: 10/07/2007].

MyCoRe se compone de un conjunto de módulos que proporcionan las funciones básicas que podrían ser requeridas en una implementación de un repositorio, incluyendo: búsqueda distribuida sobre repositorios MyCoRe dispersos geográficamente, funcionalidad OAI, soporte integrado para transmisión en línea de audio y video, gestión de ficheros, y editores de metadatos en línea. Además, ofrece un paquete para la implementación básica de un servidor de publicaciones, DocPortal. Las implementaciones locales pueden adaptar este conjunto básico para servir a sus requisitos particulares. Al contrario que MILESS (su precedente), que proporcionaba un modelo de datos estricto basado en Dublin Core cualificado, el modelo de datos de MyCoRe es completamente configurable.

MyCoRe emplea tecnología web como Java, XML o XSL. No está restringido a una base de datos particular, sino que soporta varios sistemas de bases de datos: de tipo comercial como IBM Content Manager, y software libre como MySQL y eXist XML:DB. Los componentes principales de MyCoRe son: aplicación de metadatos del documento y de las personas, sistema de ficheros internos, sistema de clasificación jerárquica, módulo WCMS (Web Content Management System) integrado, funciones de gestión de flujo y carga de trabajo; interfaces de usuario y administrador; y búsqueda distribuida e interfaces mediante OAI y Z39.50.

DocPortal
MyCoRe Demo

Navigation: DocPortal - The MyCoRe-Sample > Retrieval > for documents extended

Search mask for documents

Spezielle Suche nach Dokumenten:

Title:

Author:

Identifier:

Origin: (please select)

Subject: (please select)

Format: (please select)

Language: (please select)

Type: (please select)

Description:

Date:

Keywords:

Source:

Coverage:

Text search

in the metadata:

in the files:

Search at: ☒ this server ☐ all servers ☐ this and selected servers:

☐ Local via Webservice ☐ DocPortal Sample Server

☐ DOL of the University of Leipzig

Maximal 100 hits, 10 per page

Figura 8.6. DocPortal, *demo* del sistema MyCoRe, interfaz de búsqueda extendida de documentos [<http://mycoresamplelinux.dl.uni-leipzig.de/>]

El sistema MyCoRe está siendo utilizado en distintas universidades de Alemania (Hamburgo, Bonn, Friburgo...) aunque como aún está en desarrollo, estas aplicaciones se basan en el código de MIDESS con adaptaciones a los requisitos locales. Está especialmente dirigido a repositorios de contenido multimedia y bibliotecas digitales, complementando las funcionalidades del sistema OPUS, utilizado de forma generalizada en este país para repositorios de publicaciones académicas.

8.2.7. ARNO

El proyecto holandés ARNO (*Academic Research in the Netherlands On line*) comenzó en septiembre del 2000 a través de una colaboración de la Universidad de Tilburg, la Universidad de Twente y la Universidad de Ámsterdam. Además de diversas investigaciones de interés sobre el tema de la comunicación académica digital, uno de los resultados más importantes de este proyecto es el software del mismo nombre, ARNO, que cuenta con una media docena de implementaciones, todas en universidades holandesas y dentro de la red de repositorios institucionales DAREnet³¹.



Figura 8.7. Captura de pantalla del sistema ARNO, selección del repositorio³².

³¹ El programa Digital Academic Repositories (DARE) es desarrollado conjuntamente por las universidades holandesas, la Netherlands Organisation for Scientific Research (NWO) y la Royal Netherlands Academy of Arts and Sciences (KNAW), con el objetivo de crear repositorios institucionales. Estos repositorios utilizan principalmente los sistemas software holandeses ARNO e i-Tor.

³² Véase la sección de capturas de pantalla del sistema en el sitio Web del software: *Editor Screenshoots*: <<http://arno.uvt.nl/~arno/site/editorss.html>>. [Fecha de consulta: 10/07/2007].

Actualmente en su tercera versión, ARNO 1.2, su nota más peculiar es que no ofrece interfaz de usuario, tan sólo de administrador, y depende de otros sistemas a modo de portal como I-Tor o i-Port³³ para poder recolectar y distribuir el contenido de los repositorios ARNO.

En cuanto a las características técnicas y funcionalidades más importantes del sistema, hay que señalar que ARNO está escrito en Perl y requiere Oracle 8i, o una base de datos relacional mayor. ARNO funciona sobre los sistemas operativos Linux o Unix, y utiliza el servidor web Apache. El esquema de metadatos básico que emplea es Dublin Core, y mediante la utilización de hojas de estilo XSLT se puede aplicar otros modelos de metadatos manteniéndose compatible con OAI. Además, ARNO es neutro en cuanto a formatos (acepta cualquiera) y permite crear un registro de formato de ficheros aceptados para cada archivo. Cabe resaltar que el software cuenta con un sistema de identificación propio, llamado ARNOdocurl, para crear URL's persistentes para cada fichero de documento y página índice.

8.2.8. DiVA

DiVA (<http://www.diva-portal.org/>), el Archivo Digital Científico (Digital Scientific Archive) es un sistema de publicación electrónica resultado del proyecto del mismo nombre llevado a cabo por diversas universidades escandinavas³⁴ con el objetivo de ofrecer servicios de publicación y soluciones técnicas para repositorios locales. El proyecto comenzó en 2000 y fue completamente operativo desde enero de 2003. El sistema fue creado por el Centro de Publicación Electrónica de la Biblioteca de la Universidad de Uppsala (Suecia). Actualmente lo emplean 15 universidades de Suecia y Noruega, y todas ellas cooperan en el desarrollo del sistema.

DiVa gestiona, publica y archiva documentos a texto completo de las universidades participantes, incluyendo tesis doctorales y de grado, informes de investigación, monografías y capítulos de libro, así como *preprints* y *postprints* de artículos científicos. El sistema está basado en tecnologías Java y XML, proporciona distintos servicios de metadatos como recolección mediante el protocolo OAI-PMH, la generación de registros de catalográficos para bibliotecas (inclusión en la unión nacional de catálogos de biblioteca en Suecia, LIBRIS), o su diseminación a través de canales RSS.

³³ iPort <http://www.oclc-pica.org/content/11/pdf/iport_uk.pdf> es un sistema de portal diseñado por OCLC-PICA que proporciona acceso integrado y personalizado –a través de una única interfaz– a recursos de información y del conocimiento locales y remotos. [Fecha de consulta: 10/07/2007].

³⁴ Este consorcio fue fundado en 2002 y, además de la Universidad de Uppsala, incluye las bibliotecas de las universidades suecas de Estocolmo, Umeå, Örebro y Södertörn. En 2003, se incorporó el primer miembro de otro país, la biblioteca del estado y de la Universidad de Århus en Dinamarca. La cooperación de DiVA está abierta a todas las universidades y centros de investigación con financiación pública de Suecia y del exterior.

DiVA permite la publicación de materiales tanto electrónicos como impresos, manteniendo la coherencia entre ambas instancias de un mismo documento. Los metadatos y si es posible, el contenido del documento, son almacenados en un formato XML uniforme, el DiVA Document Format (DDF), y los metadatos pueden ser fácilmente mapeados a otros formatos como Dublin Core MARC, METS, MODS, Endnote, Reference Manager o TEI. El formato DDF se define con un esquema XML, cuenta con más de 100 elementos tanto de metadatos como de marcado estructural.

DiVA cuenta con una herramienta de administración específica para la gestión del contenido, DiVA Manager, que puede utilizarse directamente mediante un navegador web. Esta herramienta permite la carga de los documentos, pudiendo tratar en procesos separados los metadatos, resúmenes o textos completos.

Dispone de un interfaz único, llamado Portal DiVA, que actúa como un proveedor de servicios, recogiendo únicamente datos publicados por los repositorios que utilizan DiVA. Esto no impide que los datos de estos repositorios también puedan ser recolectados por otros proveedores de servicios.

El interfaz de búsqueda del sistema está construido con tecnología XML y se acompaña de una versión en texto plano que cumple las especificaciones WAI de accesibilidad. Es configurable a múltiples idiomas, aunque de momento sólo está disponible en sueco e inglés. El motor de búsqueda es Apache Lucene, y permite búsquedas estructuradas por campos o texto libre, soportando la búsqueda a texto completo.

Academic Archive On-line

Start Page/Search | **Freetext Searching** | Full-text Search

In DiVA you can find theses, dissertations and other publications in full-text from a number of Nordic universities. The publications are stored in pdf which requires [Acrobat Reader](#) to open and print them. DiVA has been developed by the Electronic Publishing Centre at Uppsala University Library.

University: ☒ Gävle (71) ☒ Gymnastik- och Idrottshögskolan (157) ☒ Jönköping (642) ☒ Karlstad (686) ☒ KTH (1853) ☒ Linköping (4052) ☒ Mittuniversitetet (31) ☒ Mälardalen (73) ☒ NTNU (1028) ☒ Skövde (27) ☒ Stockholm (1551) ☒ Södertörn (996) ☒ Umeå (855) ☒ Uppsala (2641) ☒ Växjö (768) ☒ Örebro (414)

Publication type: ☒ Articles (90) ☒ Theses (5688) ☒ Master theses / Undergraduate theses (9305) ☒ Reports (738) ☒ Chapters (5) ☒ Other (19)

Select one or more fields (automatically AND between fields)

Author: (last name) (first name)

Title: (or a word from the title)

Series: (number)

Year: (YYYY)

ISBN: (include hyphens)

Sort by: Author Ascending 20 (number of displayed hits per page)

Go to: [Freetext Searching](#) (including abstracts and keywords) | [Search Tips](#)

Search only in the database for: [Gävle](#) [Gymnastik- och Idrottshögskolan](#) [Jönköping](#) [Karlstad](#) [KTH](#) [Linköping](#) [Mittuniversitetet](#) [Mälardalen](#) [NTNU](#) [Skövde](#) [Stockholm](#) [Södertörn](#) [Umeå](#) [Uppsala](#) [Växjö](#) [Örebro](#).

These databases also include publications which are not available in full-text.

© 2000-2007 Uppsala University Library | Electronic Publishing Centre | Contact | Cookies

Figura 8.8. Página principal del Portal DiVA [<http://www.diva-portal.org/>]

El sistema ha tomado muy en cuenta la preservación a largo plazo, por lo que se asignan automáticamente identificadores persistentes a cada ítem, y estos se almacenan tanto en un repositorio local como en un archivo local. Cada publicación tiene un identificador único URN:NBN, un número bibliográfico nacional en la forma de un nombre uniforme del recurso asignado por la Biblioteca Real de Suecia. Una vez asignado el número, una copia del documento se almacena en el archivo de DiVA, y la otra se envía como un paquete de información a la Biblioteca Real para su preservación a largo plazo.

Algunas actividades de desarrollo³⁵ son la definición de metadatos de derechos y de preservación; la mejora y flexibilidad de los flujos de remisión de documentos, sistemas de búsqueda sofisticada y navegación a nivel de documento, módulo avanzado para la impresión bajo demanda, archivado de documentos a texto completo en XML y herramientas para el marcado semántico.

8.2.9. Archimede

El software Archimède (<http://archimede.bibl.ulaval.ca/>) ha sido desarrollado por la Biblioteca de la Universidad de Laval (Quebec, Canadá), con el objetivo de albergar los *preprints*, *postprints* y otras publicaciones científicas electrónicas del profesorado y personal investigador de la institución. Archimède complementa dos componentes del sistema de repositorio institucional previamente implementados por Laval: el primero gestiona las tesis y disertaciones electrónicas de la universidad (<http://www.theses.ulaval.ca/>), y el segundo proporciona una plataforma de producción para revistas y monografías electrónicas de carácter científico, disponibles a través del portal Érudit (<http://www.erudit.org>)³⁶.

Inspirado en el modelo DSpace, Archimède es un sistema descentralizado que organiza el contenido en torno a una red de comunidades de investigación auto-gestionadas. Las características principales de Archimède se plasman en su carácter multilingüe con fines de internacionalización (el texto del sistema de interfaz de usuario es independiente del código del software gracias al estándar de código abierto i18n para la codificación en UTF8), sus opciones de búsqueda mediante un motor y sistema de indexación de desarrollo propio llamado LIUS (Lucene Index Update and Search), y su flexibilidad de implementación al estar desarrollado con tecnologías Java de código abierto.

³⁵ Estos futuros desarrollos se recogen en: Uwe Klosa y Eva Müller. "The DiVA System: Current Status and Ongoing Development". En *XII Inforum, Praga, 23-25 mayo 2006*. Disponible en: <<http://www.inforum.cz/inforum2006/english/prispevek.php?prispevek=43>> [Fecha de consulta: 10/07/2007].

³⁶ Érudit <<http://www.erudit.org>> es un proyecto conjunto de la DGTIC de la Universidad de Montreal, la Biblioteca de la Universidad Laval, el servicio de búsqueda de la Universidad del Québec a Montreal y del servicio de publicaciones de la Universidad de Montreal y la Universidad Laval. Los documentos albergados en la Universidad Laval en ARCHIMEDE y en el servidor de TESIS se encuentran también enlazados con Érudit. [Fecha de consulta: 10/07/2007].



Figura 8.9. Página de inicio de Archimède en la Universidad de Laval

<http://archimede.bibl.ulaval.ca>

Archimède no depende de una plataforma específica, pudiendo funcionar sobre varios sistemas operativos (Windows, Linux, etc.) y emplear varias bases de datos relacionales compatibles con JDBC (Java Database Connectivity). Esto permite a la institución que desee utilizar Archimède una gran flexibilidad para instalar el software en una infraestructura técnica previa, aumentando la eficiencia del proceso de implementación.

En octubre de 2006 se lanzó una nueva versión del software, denominada Archimède 2.0.0, disponible para su descarga en Sourceforge (<http://sourceforge.net/projects/archimede/>), con novedades relativas a la seguridad, soporte a versiones, flujos de trabajo personalizados, formatos de metadatos adaptados, y la adopción de estándares como Java Content Repository Standard (JSR-170), y WebDAV (Developing Authoring and Versioning), que proporciona las utilidades de Windows a la Web, mediante el interfaz DeltaV. De esta manera se permiten las acciones de crear, cortar, copiar, pegar y borrar sobre los documentos y las colecciones del repositorio.

8.2.10. ETD-db

ETD-database o ETD-db (<http://scholar.lib.vt.edu/ETD-db>) es un proyecto de Virginia Tech en colaboración con la Networked Digital Library of Theses and Dissertations (NDLTD). Este proyecto ha desarrollado un software propio, ETD-db, que es un depósito de documentos basado en un conjunto de páginas web y guiones Perl que interactúan con una base de datos MySQL. Los guiones proporcionan un interfaz estándar para internautas, investigadores, autores, personal de universidades y de biblioteca, para que remitan y gestionen ficheros y metadatos relacionados con una colección de tesis y disertaciones electrónicas.

La última versión de ETD-db es la v1.8, lanzada en marzo del 2004. Su descarga está disponible de forma gratuita para los miembros de la NDLTD, acompañada de diversos manuales para su instalación, mantenimiento y actualización. Para ello precisa de un servidor en activo, preferentemente con un sistema operativo tipo Unix, en el que se deberá instalar una base de datos MySQL y un servidor web Apache. Además de la instalación estándar de Perl, precisa de algunos módulos específicos como CGI.pm, DBI, DBD:Mysql y Tie-IxHash.

Para cumplir con la versión 2.0 del protocolo OAI-PMH, precisa de un módulo adicional también disponible para su libre descarga, que incorpora utilidades para la transformación del esquema de metadatos específicos definidos por la NDLTD a los metadatos que define el protocolo, es decir, Dublin Core sin cualificar.

8.2.11. DoKS

DoKS (Document and Knowledge Sharing application) (<http://www.doks.be/>) es un proyecto que trata de proporcionar una herramienta que facilite el almacenamiento, intercambio y recuperación de documentos (e información en general) a través de Internet. El objetivo inicial de este proyecto se centra en la recuperación de las tesis de los institutos de educación superior de la región belga de Flandes, aunque en un futuro se pretende que sirva también a la gestión del conocimiento, gestión documental y e-learning.

DoKS, formalmente conocido como el proyecto “Biblioteca Digital contenedora de las Tesis de los Institutos Flamencos de Educación Superior”, es una iniciativa de la biblioteca de la Escuela Superior Católica Kempen (Katholieke Hogeschool Kempen), en colaboración con la Universidad Católica de Lovaina (KULeuven) y otros socios, y está financiada por el Instituto Flamenco para la Promoción de la Innovación a través de la Ciencia y la Tecnología (IWT).

La arquitectura del software se basa en un conjunto de estructuras Java de código abierto entre las que se encuentra OAlcat para la implementación del protocolo OAI-PMH (véase *epígrafe* 8.3.1.4). DOKs funciona sobre cualquier plataforma con Java, pudiéndose utilizar sobre Windows o Linux. Es necesario instalar los paquetes Java, un

sistema de gestión de bases de datos, Apache Ant y el servidor web Tomcat. Es independiente de la base de datos, por lo que puede funcionar sobre MySQL, Microsoft SQL Server, Oracle, PostgreSQL, etc.

Esta herramienta se construye sobre un sistema de registros (*records*) y carpetas (*folders*). Una pieza de información en DoKs se denomina registro, y cada registro consiste en un número de campos definidos (p.ej.: una FAQ puede tener un campo de pregunta y otro de respuesta). Estos campos son definidos en la estructura del registro, y el sistema permite la definición de estructuras propias con campos arbitrarios. Estos registros se pueden organizar en una jerarquía de carpetas de forma similar a los directorios en un sistema de ficheros, de manera que cada carpeta puede contener registros o subcarpetas. DoKs contiene un módulo de guiones similar a JavaScript, BeanShell, que hace a los ficheros y carpetas extensibles y configurables, y cuenta con un mecanismo de permisos para controlar el acceso a registros y carpetas, lo que implica que DoKs permite definir usuarios y grupos de usuarios.

DoKs cuenta con un conjunto de funcionalidades implementadas comunes a otros sistemas de archivos abiertos, como la configuración de múltiples conjuntos de metadatos, interfaz web multilingüe y configurable, la importación y exportación en XML, exportación de metadatos en formato MARCXML, o la búsqueda por texto libre y por campos de metadatos³⁷.

8.2.12. Kepler

Kepler (<http://kepler-project.org/>) es un proyecto de colaboración para desarrollar un sistema de código abierto de flujo de trabajo científico, basado en el sistema Ptolemy II de la Universidad de California. Su finalidad principal consiste en que los investigadores puedan colgar sus documentos de trabajo en él, y a su vez puedan acceder a la documentación producida y publicada por otros investigadores. Es aplicable a diversas disciplinas desde biología, ecología, astronomía a subdisciplinas de la informática, como la bioinformática, ecoinformática o geoinformática.

Actualmente no existe una versión completa de Kepler, pero sí una versión beta de prueba de enero de 2007, Kepler 1.0.0.beta3, que está disponible para Windows, MacOS y Linux. Está desarrollado con tecnología Java por lo que precisa Java 1.4 para su instalación.

Kepler permite que cada investigador se convierta en un proveedor de datos OAI [Silió, 2005]. Contiene un “archivelet”³⁸ autoinstalable para autoarchivar publicaciones.

³⁷ Una lista de las principales funcionalidades implementadas y planificadas en DoKs se recoge en: <http://doks.khk.be/do/folder/View?dispatch=info&folderId=Fold413ebf17fb3fd0f300fb3fd0f9ee000a> [Fecha de consulta: 10/07/2007].

³⁸ Un “archivelet” es un archivo de un trabajo académico individual, opuesto al archivo completo de toda la institución o disciplina. Es un término creado por el proyecto Kepler. Véase el *Glosario del Movimiento por el Acceso Abierto*: <http://www.infodoctor.org/bibliotecapublicadelaciencia/glosario.php>, traducción al castellano de la *Guide to the Open Access Movement* <http://www.earlham.edu/~peters/fos/guide.htm> de Peter Suber. [Fecha de consulta: 10/07/2007].

Pretende integrar las bibliotecas digitales y los grupos de investigación, que así pueden tener un control constante sobre su biblioteca digital. Contiene dos posibilidades, una enfocada a grandes comunidades y otra para edición individual. En el caso de las comunidades se distribuye un software a cada miembro, un archivelet; después el servidor reúne toda la información distribuida en archivelets y realiza las búsquedas. En el caso de los editores individuales que quieran instalar un repositorio que cumpla con OAI, existen unos archivelets genéricos.

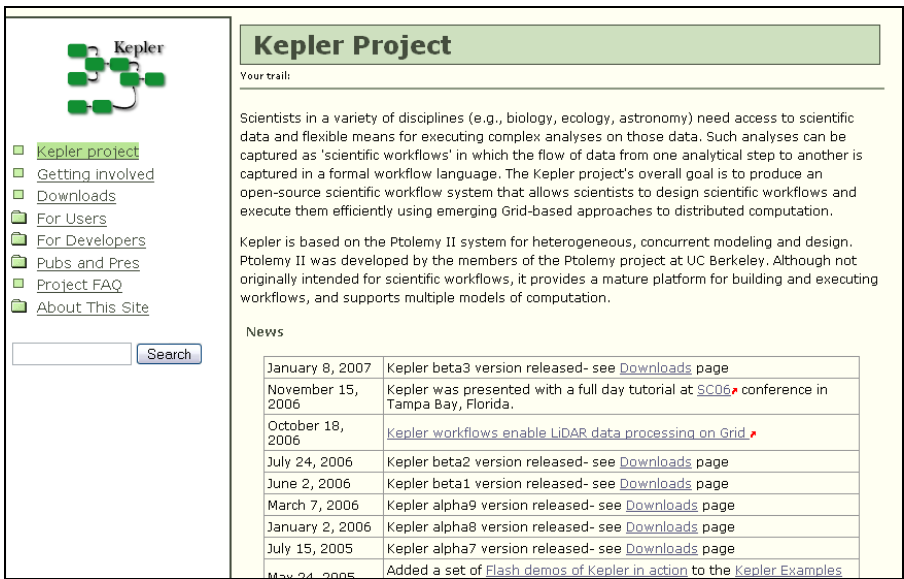


Figura 8.10. Página principal del proyecto Kepler [<http://kepler-project.org/>]

Otras herramientas similares a Kepler son PDLib³⁹ y el software de FreeScience⁴⁰. El primero es un software para construir colecciones personales digitales (Personal Digital Libraries) conforme a OAI, desarrollado en el Instituto Tecnológico de Monterrey, México. El cuanto a FreeScience, es un proyecto de digitAlexandria (<http://www.bdaweb.net/>), que pretende que todos los investigadores y estudiantes puedan compartir sus publicaciones y otros resultados de su investigación mediante una red bibliotecaria P2P que además cumple el protocolo OAI-PMH. Su software cliente es de libre descarga y se puede instalar en todo tipo de sistemas operativos.

³⁹ PDLib <<http://copernico.mty.itesm.mx/pdlib/>> se ha descrito con mayor detalle en el capítulo 6, dedicado al movimiento OAI en Latinoamérica. [Fecha de consulta: 10/07/2007].

⁴⁰ Freescience <<http://www.bdaweb.net/template/inc/freescience/>> un proyecto de digitAlexandria. [Fecha de consulta: 10/07/2007].

8.3. APLICACIONES GENERADORAS DE PROVEEDORES DE DATOS

Las estrategias de establecimiento de proveedores de datos requieren un trabajo coordinado por los profesionales de la biblioteca, y en general, no implican un esfuerzo excesivamente gravoso para la mayoría de las organizaciones al tiempo que aportan considerables beneficios por su capacidad de intercambio de metadatos mediante un protocolo estandarizado. La implementación de OAI es a menudo interpretada erróneamente como un importante requerimiento de desarrollo de sistemas; sin embargo, no siempre es necesario realizar grandes modificaciones en las infraestructuras de bibliotecas digitales ya existentes, ya que a menudo las implementaciones OAI son un mero parche en el sistema que requiere la instalación de unos pocos guiones de programación o consisten simplemente en añadir un nuevo módulo al software existente.

En realidad, la tarea más compleja de convertirse en un proveedor de datos es la de establecer un vínculo con su propia colección. El proveedor de datos debe ser capaz de buscar a través de ella y recuperar las descripciones de registros que suministrar al recolector cuando éste acuda a él en busca de registros de metadatos, tanto si es la primera vez y precisa de todos los registros que integran la colección como si es una actualización y tan sólo necesita las nuevas incorporaciones.

Hoy en día existen diversas soluciones efectivas, simples y rápidas para implementar OAI en infraestructuras existentes y convertirlas en proveedores de datos conformes al protocolo OAI-PMH [Halbert y Milewicz, 2005]. Algunas de las infraestructuras de partida con las que se permite conexión directa son: bases de datos relacionales de una o múltiples tablas, sistemas de ficheros en XML o pasarelas Z39.50.

En los siguientes epígrafes se describen algunas de las herramientas disponibles capaces de generar fácilmente proveedores de datos a partir de colecciones de documentos electrónicos ya existentes, organizadas en razón de la infraestructura previa de la colección. Los lenguajes de programación empleados para desarrollar estas aplicaciones son en su mayoría PHP, Java o Perl; así como tecnologías web ASP o JSP.

Entre los desarrolladores de las aplicaciones seleccionadas merece especial mención el Laboratorio de Investigación en Bibliotecas Digitales (DLRL) del Instituto Politécnico de Virginia (Virginia Tech), el proyecto OAI-PMH de la Biblioteca de la Universidad de Illinois en Urbana Champaign, así como el Laboratorio de Tecnologías Interactivas y Cooperativas (ICT) y el Centro de investigación en Tecnologías de Información y Automatización (CENTIA), de la Universidad de Las Américas Puebla, México. Estos grupos de investigación han creado varias herramientas generadoras de proveedores de datos, junto a otras herramientas que soportan otros aspectos del protocolo OAI-PMH. También se recogen herramientas desarrolladas por universidades, como las estadounidenses Old Dominion y Chapel Hill (en colaboración con el proyecto ibiblio), o la Universidad de Oldenburg (Alemania); el departamento de Investigación de Online Computer Library Center (OCLC Research) e incluso de empresas como ZZ/OSS Information Networking.

8.3.1. Bases de datos relacionales

En este epígrafe se recogen un total de ocho aplicaciones que generan proveedores de datos a partir de colecciones de metadatos almacenadas en distintos tipos de bases de datos relacionales, y que son: OAIB (Open Archives in a BOX), OAlbiblio, VTOAI (Virginia Tech OAI), OAICat, VOA, el proveedor de datos OAI en PHP de la Universidad de Oldenburg (Alemania) y las dos aplicaciones proveedores de datos ASP y JSP de la Universidad de Illinois. Para el conocimiento de estas herramientas ha resultado de gran ayuda el trabajo previo realizado por Villegas Ortega para su tesis doctoral, y que se recoge en su capítulo 2 [Villegas Ortega, 2005].

8.3.1.1. OAIB (*Open Archives in a Box*)

OAIB (Open Archives in a Box) (<http://dlt.ncsa.uiuc.edu/oaib/>) es una sencilla aplicación para exportar metadatos almacenados en un sistema de gestión de bases de datos relacionales como Microsoft Access, Oracle o MySQL y crear así un proveedor de datos conforme al protocolo OAI-PMH. Se creó en 2001 por DLT (Digital Libraries Technologies), grupo de tecnologías para bibliotecas digitales del NCSA (National Center for Supercomputing Applications)⁴¹, en Estados Unidos.

OAIB ha sido desarrollada en Java y puede ser instalada en plataformas Windows, Linux y MacOS. Se basa en una estructura denominada *Components for Constructing Open Archives* (COCOA) que se incluye en el paquete de distribución, y precisa de un servidor Java como Apache Tomcat. La última versión del programa es la OAIB 2.0, y fue publicada en septiembre de 2005.

La herramienta es fácil de instalar sin necesidad de conocimientos de programación, ya que no es necesario hacer ningún tipo de implementación por parte del usuario. Al configurar el sistema se crea un fichero de configuración en XML que establece cómo acceder a la base de datos y como traducir los “verbos” OAI en consultas SQL, así como traducir las respuestas a estas consultas en SQL en registros de metadatos que se entregan mediante un cliente OAI.

El inconveniente de OAIB es que es una herramienta orientada a colecciones descritas en una sola tabla relacional, y en el caso de que la colección que se desea compartir se encuentre en una base de datos relacional en tablas múltiples, obligará al administrador a realizar el volcado de los metadatos necesarios a una sola tabla, y generar de esta forma el servidor OAI.

⁴¹ El grupo DLT (Digital Libraries Technologies) <<http://dlt.ncsa.uiuc.edu/>> forma parte del NCSA (National Center for Supercomputing Applications), un centro nacional de investigación supercomputación dependiente de la National Science Foundation (Estados Unidos) y con sede en la Universidad de Illinois en Urbana-Champaign.

8.3.1.2. OAIbiblio PHP Data Provider

OAIbiblio (<http://www.ibiblio.org/oaibiblio/>) es una aplicación en PHP que permite la generación de servidores OAI para colecciones descritas en múltiples tablas bajo un Sistema de Gestión de Bases de Datos Relacionales en MySQL. OAIbiblio es un desarrollo de *ibiblio: The Public's Library and Digital Archive*, una colección de colecciones de Internet elaborada en colaboración por el Center for the Public Domain y la Universidad de Carolina del Norte en Chapel Hill.

Algunas de sus funcionalidades destacadas son: mapeo de metadatos empleando ficheros XSL (eXtensible Stylesheet Language) independientes, retención de metadatos de almacenamiento con tecnología WDDX (Web Development Data Exchange), asociaciones únicas de documentos con formatos y conjuntos de metadatos, nivel de almacenamiento separado de la aplicación y de las consultas (facilitando así la instalación y modificación del sistema), y soporte a estructuras de bases de datos con múltiples tablas.

La dificultad de la herramienta radica en que, además de requerir la instalación de PHP y el uso de DOM (Document Object Model) para administrar los objetos XML, es necesario que el usuario configure manualmente algunos componentes generados antes de que el servidor OAI pueda ser utilizado.

Su principal inconveniente es que no se actualiza desde marzo de 2004, y se ha quedado en una versión beta simple que tan sólo expone metadatos pero que no permite su gestión (edición, creación, borrado, etc.)

8.3.1.3. VTOAI (Virginia Tech OAI)

VTOAI (Virginia Tech OAI) o OAI-PMH2 PERL Implementation⁴² es una aplicación en Perl enfocada a la generación del esqueleto o estructura principal de un servidor OAI a partir de los metadatos almacenados en bases de datos relacionales multitablas, y que se ajuste a la versión 2.0 del protocolo OAI-PMH. Es un producto del laboratorio de investigación en bibliotecas digitales de Virginia Tech, el DLRL (Digital Library Research Laboratory), y no se actualiza desde junio de 2002.

Es una herramienta que no requiere instalación ni compilación, ya que sólo es necesario copiar los guiones Perl y contar con un servidor web con capacidades para utilizar guiones CGI. Las extensiones, configuraciones y contenedores se especifican empleando XML Schema, y cuenta con composición de código para componentes o bibliotecas de componentes separados.

⁴² Información sobre VTOAI en: <<http://www.dlib.vt.edu/projects/OAI/software/vtoai/vtoai.html>>. [Fecha de consulta: 10/07/2007].

8.3.1.4. OAI/Cat

OAI/Cat⁴³ es una aplicación en Java que proporciona una estructura de repositorio conforme al protocolo OAI-PMH v2.0. Este entorno puede configurarse para trabajar con cualquier repositorio de datos implementando ciertos interfaces Java. Algunos de estos interfaces se incluyen en el paquete de distribución a modo de ejemplo. OAI/Cat es un producto de OCLC, quien lo distribuye con una licencia propia (OCLC Research Public License), y desde 2006 también mediante licencia Apache, versión 2.

Esta herramienta es muy similar a VTOAI en el sentido de que genera el esqueleto o estructura principal de un servidor OAI a partir de bases de datos relacionales simples o multitablas. OAI/Cat también exige trabajo adicional al usuario, pues es necesario especificar parámetros en un archivo de configuración, un proceso que podría resultar complejo para un usuario administrador que solamente conoce la estructura de la colección que desea compartir.

8.3.1.5. VOAI

VOAI⁴⁴ es un generador automático de servidores de metadatos bajo el protocolo OAI-PMH, creado en 2005 por el Laboratorio de Tecnologías Interactivas y Cooperativas (ICT) y el Centro de investigación en Tecnologías de Información y Automatización (CENTIA), de la Universidad de Las Américas Puebla, México.

Los desarrolladores de la herramienta estudiaron cinco generadores automáticos de servidores OAI (los cuatro presentados en los apartados anteriores junto a RVOT, recogido en el apartado 8.3.2.5) y trataron de subsanar las deficiencias que presentaban, como el soporte a bases de datos multitablas, las necesidades de modificación de la estructura de la base de datos o los requisitos de implementación por parte del usuario. Con todas estas mejoras, VOAI permite la generación automática de servidores OAI para colecciones almacenadas en cualquier base de datos relacional, sin obligar a tener almacenados todos los metadatos en una sola tabla, y dando libertad al usuario de indicar las consultas en SQL que se utilizarán para recuperar los metadatos. Aleja al usuario de cualquier detalle de implementación quién únicamente tiene que colocar el código generado en un contenedor de servlets o utilizar la herramienta de instalación que se proporciona para el sistema operativo Microsoft Windows. Puede ser utilizado en cualquier plataforma, ya que emplea Java como lenguaje de programación.

Aunque ya no se encuentran accesibles, cabe mencionar una aplicación de VOAI para la generación automatizada de dos proveedores de datos de la Universidad de

⁴³ Información sobre OAI/Cat <<http://www.oclc.org/research/software/oai/cat.htm>> en la página web de la OCLC. [Fecha de consulta: 10/07/2007].

⁴⁴ Descripción de VOAI, en: <http://ict.udlap.mx/software/voai_dist/descripcion.html>. [Fecha de consulta: 10/07/2007].

Valladolid⁴⁵: uno que proporcionaba acceso a la colección de Manuscritos de la Biblioteca de Santa Cruz, y otro está destinado a la colección digital de Incunables de la biblioteca universitaria. Ninguno de estos dos proveedores constituyen un repositorio institucional, ni ofrecen interfaz de usuario con servicios de búsqueda o remisión de documentos, sino tan sólo los metadatos OAI para su recolección mediante consultas OAI-PMH.

8.3.1.6. *PHP OAI Data Provider*

“PHP OAI Data Provider” (<http://physnet.uni-oldenburg.de/oai/>) es un desarrollo de la Universidad de Oldenburg, Alemania. Cumple completamente con el protocolo OAI-PMH 2.0, incluyendo la compresión de respuestas en XML lo que reduce significativamente el volumen de datos a transferir.

Es una aplicación de configuración sencilla y se puede conectar con varias bases de datos existentes empleando la capa de abstracción de bases de datos en PEAR⁴⁶. Puede instalarse en plataformas Linux, Unix y Windows, requiere tecnologías Apache y PHP y soporta bases de datos relacionales como Oracle o MySQL. Su última versión, PHP OAI 2-1.8.0, es de junio de 2005.

Este proveedor de datos es utilizado por el repositorio de partituras musicales de la Universidad de Indiana (Indiana University Sheet Music)⁴⁷ así como 'IFEApub'⁴⁸, repositorio de publicaciones del Instituto Francés de Estudios Andinos.

8.3.1.7. *Aplicaciones de la UIUC para colecciones en bases de datos*

La Universidad de Illinois en Urbana-Champaign (UIUC) cuenta con dos herramientas para crear proveedores de datos OAI a partir de bases de datos relacionales, compatibles con la versión 2.0 del protocolo OAI-PMH e instalables sobre plataformas Windows. La primera ha sido desarrollada con Microsoft Active Server Pages (ASP) y diseñada para funcionar bajo un servidor web Microsoft IIS, mientras que la segunda utiliza tecnología JSP (Java Server Pages). Ambas están disponibles para su descarga en Sourceforge junto al resto de herramientas OAI creadas por la UIUC⁴⁹.

⁴⁵ Las URL base de los proveedores de datos OAI de la Universidad de Valladolid eran <<http://gutenberg.dcs.fi.uva.es:8080/uvaoui/proveedor>> para la colección de manuscritos y <http://gutenberg.dcs.fi.uva.es:8080/OAI_UVA/Oai_incunables> para la de Incunables. Aunque ambas han dejado de funcionar, aún figuran en el registro oficial de proveedores de datos de la OAI <<http://www.openarchives.org/Register/BrowseSites>>. [Fecha de consulta: 10/07/2007].

⁴⁶ Para más información sobre la PEAR Database Abstraction Layer, véase: <<http://pear.php.net/package/DB>>. [Fecha de consulta: 10/07/2007].

⁴⁷ Repositorio de partituras musicales de la Universidad de Indiana (Indiana University Sheet Music): <<http://www.let.rs.indiana.edu/s/sheetmusic/>>. [Fecha de consulta: 10/07/2007].

⁴⁸ Repositorio de publicaciones del Instituto Francés de Estudios Andinos (IFEA), en: <<http://www.ifeanet.org/>>

⁴⁹ Breve descripción sobre las herramientas desarrolladas por la UIUC y enlaces para su descarga en Sourceforge: <<http://uilib-oai.sourceforge.net/>>. [Fecha de consulta: 10/07/2007].

“ASP OAI 2.0 Data Provider for Database” es una aplicación que emplea una base de datos Access para almacenar la información administrativa sobre los registros de metadatos así como los propios metadatos. Los usuarios pueden incluir o crear sus propios gestores de metadatos para diferentes formatos. En la implementación de ejemplo se incluyen dos gestores: uno que lee los registros de la base de datos y los formatea a Dublin Core (`metadataPrefix=oai_dc`) y el otro que los transforma en registros MARC (`metadataPrefix=marc`).

Y “JSP OAI 2.0 Data Provider for Database”, ha sido implementado en JSP y diseñado para funcionar bajo un servidor Apache / Tomcat (4.0 o superior) y Java Runtime Environment (1.4 o superior), con una arquitectura en la que los metadatos son almacenados enteramente en una base de datos MySQL. Los usuarios pueden proporcionar sus propios gestores de metadatos aunque por defecto cuenta con un gestor que lee los registros de la base de datos y los formatea a Dublin Core (`metadataPrefix=oai_dc`).

8.3.2. Sistemas de ficheros XML

En el grupo de aplicaciones generadoras de proveedores de datos OAI a partir de infraestructuras previas que consistan en metadatos almacenados en un sistema de ficheros XML, destacan herramientas como XOAI, jOAI, RVOT y las dos aplicaciones ASP de la Universidad de Illinois en Urbana-Champaign. Se ha excluido del análisis el proveedor de datos de Virginia Tech⁵⁰ por no cumplir con la versión 2.0 del protocolo OAI-PMH.

8.3.2.1. XOAI

Al igual que VOAI, la plataforma XOAI⁵¹ es un producto del laboratorio de Tecnologías Interactivas y Cooperativas (ICT) y el Centro de investigación en Tecnologías de Información y Automatización (CENTIA), de la Universidad de Las Américas Puebla, México.

Las herramientas que componen XOAI permiten que cualquier colección de documentos XML pueda ser configurada para compartir sus metadatos a través de un servidor para colecciones almacenadas en bases de datos XML nativas, por lo que lo han denominado OAI-XML. Algunas de estas herramientas son la de asignación de encabezados (*headers*) a los registros de metadatos, o la herramienta de configuración

⁵⁰ XML File-based OAI Data Provider: <<http://www.dlib.vt.edu/projects/OAI/software/oai-file/oai-file.html>>, del Digital Library Research Laboratory de Virginia Tech. [Fecha de consulta: 10/07/2007].

⁵¹ Para más información, documentación y descarga de XOAI, consúltese: <<http://ict.udlap.mx/software/xoai>>. [Fecha de consulta: 10/07/2007].

del servidor. Proporcionan una forma sencilla para configurar dichas colecciones y permiten contar con una implementación única de los servidores OAI-XML, evitando con ello la necesidad de lidiar con aspectos de configuración.

El diseño y la implementación realizados bajo el modelo MVC (Modelo-Vista-Controlador) permiten que el usuario tenga el control total sobre la presentación final de las respuestas que el servidor envía al cliente; así como del funcionamiento del propio servidor.

8.3.2.2. *jOAI Data Provider*

jOAI⁵² es una herramienta que se compone de un proveedor de datos y un recolector OAI, y ha sido desarrollada por el centro del programa DLESE de la University Corporation for Atmospheric Research (UCAR). El proveedor de datos de jOAI permite que los ficheros XML de un sistema de ficheros cualquiera sean expuestos como ítems de un repositorio OAI. Por su parte, el recolector recupera los registros de metadatos de proveedores de datos OAI remotos, los almacena en el sistema de ficheros local y los comprime en un paquete zip. Soporta distintos formatos de metadatos tanto Dublin Core sin cualificar (*oai_dc*) como otros formatos propios de la biblioteca digital DLESE (*adn*, *dlese_anno*, *dlese_collect* y *news_opps*).

jOAI está programado en Java, puede instalarse en Windows, Linux, Mac OS o en sistemas UNIX y corre sobre un contenedor de servlets como Apache Tomcat. La última versión de jOAI es la 3.0.11 y se distribuye con licencia GNU GPL.

8.3.2.3. *PEAR::OAI*

PEAR::OAI⁵³ es una librería de clases PHP diseñada por ZZ/OSS Information Networking para establecer servidores OAI, empleando clases PEAR (PHP Extension and Application Repository)⁵⁴. El objetivo de sus desarrolladores era que la librería proporcionara un servidor (o repositorio) al mismo tiempo que un interfaz de cliente (harvester), aunque hasta ahora sólo se ha implementado el servidor OAI, y no se ha actualizado desde su versión 0.4.1, publicada en Sourceforge.net en abril de 2004.

De configuración sencilla y flexible, se compone de un sistema modular que permite la conexión con varios interfaces de almacenamiento de datos que tengan diferente estructura. Cuenta con un sistema de plantillas integrado para proporcionar

⁵² Documentación y descarga de jOAI Software en: <http://www.dlese.org/dds/services/joai_software.jsp>. [Fecha de consulta: 10/07/2007].

⁵³ PEAR::OAI: <http://www.zzoss.com/phpwiki/index.php/PearOai020b.html?nav_node=Software> [Fecha de consulta: 10/07/2007].

⁵⁴ PEAR (PHP Extension and Application Repository) <<http://pear.php.net/>> es el depósito oficial de código PHP. Sigue la misma idea de CPAN (Comprehensive Perl Archive Network) <<http://www.cpan.org/>> para los amantes de PERL. [Fecha de consulta: 10/07/2007].

metadatos en distintos formatos de salida, y permite la compresión en XML, reduciendo el volumen de datos para la transferencia.

8.3.2.4. Proveedores de datos ASP para sistemas de ficheros XML de la UIUC

La Universidad de Illinois en Urbana-Champaign (UIUC) cuenta con dos herramientas⁵⁵ programadas en ASP e instalables sobre plataformas Windows, ambas para crear proveedores de datos OAI compatibles con la versión 2.0 del protocolo OAI-PMH a partir de sistemas de ficheros XML.

El primer proveedor (*ASP OAI 2.0 Data Provider for File System*) almacena los metadatos en un sistema de ficheros y directorios mediante registros en MARC formateado en XML, mientras que la segunda (*ASP OAI 2.0 Data Provider for File/Database Hybrid*) emplea un sistema mixto: una base de datos relacional tipo Access para almacenar la información administrativa OAI sobre los registros de metadatos, mientras que los metadatos en sí mismos son almacenados en un sistema de ficheros en MARC formateado en XML que son referenciados mediante punteros que se incluyen en la información administrativa.

En ambos casos los usuarios pueden proporcionar sus propios gestores de metadatos, aunque por defecto se ofrecen tres: uno que emplea tecnología XSLT para transformar de forma dinámica los datos en XML MARC en el formato Dublin Core sin cualificar (`metadataPrefix=oai_dc`), otro que utiliza también tecnología XSLT que transforma los registros XML en registros MARC (`metadataPrefix=marc`); y un último formato que ofrece como resultado el fichero tal como es almacenado en el disco para mostrar otras posibilidades en el proceso de registros de metadatos (`metadataPrefix=marc_direct`). En este sentido, los usuarios pueden emplear cualquier tipo de tecnología de procesamiento o transformación de metadatos para formato específicos mientras se proporcionen los gestores de metadatos apropiados.

8.3.2.5. Rapid Visual OAI Tool

Rapid Visual OAI Tool (RVOT) (<http://rvot.sourceforge.net/>) es una herramienta creada por Sathish Kumar, estudiante del grupo de bibliotecas digitales de la Universidad de Old Dominion para construir gráficamente un repositorio OAI-PMH a partir de una serie de archivos que contengan registros en cualquier formato aceptado, como Marc. RVOT permite pasar visualmente los registros del formato nativo al formato `oai_dc` y una vez que esto se ha llevado a cabo, es capaz de responder a las solicitudes OAI-PMH.

⁵⁵ Breve descripción sobre las herramientas desarrolladas por la UIUC y enlaces para su descarga en Sourceforge: <<http://uilib-oai.sourceforge.net/>>. [Fecha de consulta: 10/07/2007].

El sistema está programado en Java y viene acompañado de un servidor HTTP y un sistema de gestión de OAI-PMH. Disponible para Windows y Linux / Unix, su última versión es la 1.0, que se publicó en noviembre de 2002 con licencia de código abierto de la Universidad de Illinois y el NCSA.

Debido a que los metadatos de los registros de la colección no están disponibles en bases de datos relacionales, es necesario especificar en RVOT todos y cada uno de los registros de la colección. De esta manera, aunque representa una excelente oportunidad para poder compartir cualquier archivo almacenado en el disco duro, RVOT tiene un problema de escalabilidad: si se desean compartir colecciones que consistan de una gran cantidad de registros (miles por ejemplo), el proceso de especificación de registros se vuelve sumamente tedioso para el usuario.

Por otro lado, aunque el diseño de RVOT permite que sea fácilmente extendido para soportar otros formatos de metadatos, es necesario que el usuario diseñe su propio parche para la importación de metadatos, lo sitúe en un directorio determinado y realice algunos cambios en diversos parámetros del fichero de código fuente.

8.3.3. Pasarelas Z39.50: ZMARCO

ZMARCO⁵⁶ es una aplicación para establecer proveedores de datos que cumplan con el protocolo OAI-PMH versión 2.0, y que, al igual que las herramientas descritas en el epígrafe anterior, ha sido desarrollada por el proyecto de recolección de metadatos de la Universidad de Illinois en Urbana-Champaign. Su última versión, la 0.2, fue publicada en junio de 2003.

La 'Z' significa Z39.50; 'Marc' significa MACHine-Readable Cataloging; y la 'O' significa OAI, Open Archives Initiative. En este sentido, la razón fundamental de ZMARCO se basa en que Z39.50 y MARC son dos estándares muy extendidos en el mundo de la biblioteca tradicional, mientras que OAI-PMH está siendo rápidamente adoptado como un protocolo ligero para el intercambio de metadatos en el seno de la comunidad de bibliotecas digitales. Por lo tanto, resulta de gran utilidad desarrollar un instrumento que permita que el complejo protocolo Z39.50 y el formato MARC puedan ser utilizados para la creación de proveedores de datos OAI, haciendo que la cantidad enorme de datos que ya están disponibles a través de estos estándares esté también disponibles mediante el protocolo OAI-PMH.

ZMARCO está programado en Visual Basic y VBScript y es sencillo de modificar si fuese necesario. Una de sus desventajas es que la transformación de los registros MARC en ocasiones resulta problemática y precisa de asistencia por parte de los administradores.

⁵⁶ ZMARCO, disponible para su descarga en Sourceforge: <<http://zmarco.sourceforge.net/>>. [Fecha de consulta: 10/07/2007].

8.4. RECOLECTORES Y PROVEEDORES DE SERVICIOS

Según su definición técnica, un proveedor de servicios lleva a cabo consultas OAI-PMH a proveedores de datos y emplea los metadatos recolectados como bases de servicios de valor añadido. De esta manera, un proveedor de servicios recolecta los metadatos expuestos por los proveedores de datos.

A continuación se describe una serie de aplicaciones capaces de generar fácilmente proveedores de servicios a partir de sus recolectores correspondientes. La mayor parte de las herramientas generadoras de proveedores de servicios y los paquetes software recolectores han sido desarrolladas por los mismos creadores de aplicaciones para proveedores de datos, como el Laboratorio de Investigación en Bibliotecas Digitales (DLRL) del Instituto Politécnico de Virginia (Virginia Tech), el proyecto OAI-PMH de la Biblioteca de la Universidad de Illinois en Urbana Champaign, el grupo de Bibliotecas Digitales de la Universidad de Old Dominion o el departamento de Investigación de la OCLC (OCLC Research). E incluso, en algunos casos se han elaborado al mismo tiempo y en el mismo paquete software un proveedor de datos y un recolector, como lo demuestra el ejemplo de las aplicaciones jOAI y PEAR::OAI que ya se han comentado en el epígrafe anterior.

En general, las herramientas recogidas en esta sección fueron desarrolladas en los comienzos de la iniciativa OAI, y en muchas ocasiones han dejado de mantenerse e incluso de distribuirse⁵⁷. El último en aparecer, y el más recientemente actualizado es el recolector del Public Knowledge Project, el cual está siendo adoptado por un número cada vez mayor de proveedores de servicios.

8.4.1. Arc

Arc es el recolector de metadatos desarrollado por y para Arc, un servicio de búsqueda federado, basado en el protocolo OAI-PMH, que permite buscar información en varios repositorios OAI desde un único interfaz de búsqueda. Incluye un recolector de metadatos que funciona con repositorios que cumplan cualquiera de las versiones del protocolo OAI-PMH, así como un motor de búsqueda con un sencillo interfaz para interrogar una base de datos relacional en la que se han almacenado previamente los metadatos recolectados de los distintos repositorios.

⁵⁷ Se han descartado herramientas como My.OAI que recientemente han dejado de mantener su sitio web y de distribuir el software recolector, que en este caso había sido desarrollado en Perl por la empresa FS Consulting, Inc., también desaparecida.

El recolector de Arc, desarrollado por el Grupo de Bibliotecas Digitales de la Universidad de Old Dominion (<http://dlib.cs.odu.edu>), se basa en tecnologías Java y requiere de un kit de desarrollo en Java (JDK 1.4), un servidor web (Tomcat 4.0x), y un servidor de bases de datos relacionales (como Oracle o MySQL). Arc puede configurarse para una comunidad específica, y para ello es posible aplicar ciertas extensiones y adaptaciones a cada comunidad. Arc es distribuido bajo licencia NCSA Open Source License y se puede descargar en Sourceforge⁵⁸, siendo su última versión, oai-arc 1.0, de octubre de 2006.

8.4.2. DP9

Al igual que Arc, DP9 (<http://dlib.cs.odu.edu/dp9/>) fue desarrollado por el Grupo de Bibliotecas Digitales de la Universidad de Old Dominion, en colaboración con el Laboratorio Nacional de Los Álamos (<http://lib-www.lanl.gov>) en 2001. DP9 es un servicio de pasarela de código abierto que permite a motores de búsqueda como Google la indexación de archivos OAI [Xiu, 2001]. Estos buscadores no son capaces de indizar los contenidos de un repositorio y no gestionan los ficheros XML con facilidad. Para resolver estas dificultades, DP9 asigna una URL persistente a los registros del repositorio, y los convierte en una consulta contra el repositorio apropiado cuando se requiere esa URL. Esto permite a los motores de búsqueda que no soportan el protocolo OAI indizar la “Web profunda” que contienen los repositorios OAI. DP9 también soporta un servicio de conversión de nombres: dado un identificador OAI, este responde con una página HTML, un fichero XML o reenvía la consulta al proveedor de datos apropiado. Se considera que si se aplica esta pasarela a otros protocolos se puede alcanzar un amplio rango de recursos interoperables en la Web.

DP9 consta de tres componentes principales: un empaquetador (wrapper) de URL, un gestor OAI y un procesador XSLT, todos ellos sobre un servidor Apache Tomcat. El empaquetador de URL acepta las URL persistentes y hace una llamada a las aplicaciones JSP internas. EL gestor OAI lleva a cabo las consultas OAI en nombre del motor de búsqueda web. El procesador XSLT transforma las respuestas en XML de los archivos OAI en un formato HTML adecuado para un motor de búsqueda. El empleo de XSLT permite a DP9 el uso de cualquier formato de metadatos en XML simplemente añadiendo un fichero XSL.

DP9 es un servicio de pasarela, ya que no almacena los registros OAI, sino que tan sólo reenvía las consultas a los proveedores correspondientes, de manera que la calidad del servicio depende de la disponibilidad de estos proveedores. Se basa en la lista de proveedores del sitio web de la Open Archives Initiative.

⁵⁸ ARC source, disponible para su descarga en Sourceforge: <<http://oaiarc.sourceforge.net/>>. [Fecha de consulta: 10/07/2007]

8.4.3. PKP Open Archives Harvester

The PKP Open Archives Harvester (<http://pkp.sfu.ca/?q=harvester>) es un recolector y proveedor de servicios creado en PHP por el proyecto Public Knowledge Project, una iniciativa conjunta de las universidades canadienses Simon Fraser y British Columbia con el apoyo de SPARC, con el objetivo general de fomentar la expansión y mejora del acceso a la investigación. Con este fin, el proyecto ha desarrollado otras dos herramientas, Open Journal Systems (OJS) y Open Conference Systems (OCS), que permiten crear proveedores de datos OAI para colecciones de revistas y conferencias respectivamente.

A partir de su versión 2, PKP Harvester permite recolectar metadatos en una variedad de esquemas, desde Dublin Core no cualificado, hasta la extensión de Dublin Core para los sistemas OJS y OCS, pasando por MODS y MARCXML. Además, permite la recolección selectiva mediante el uso de *sets* y sellos de fechas. De forma posterior a la recolección, se puede llevar a cabo un filtrado o normalización de los metadatos recopilados.

PKP Harvester presenta un interfaz de búsqueda flexible que permite búsqueda simple y avanzada en todos los archivos recolectados, ya que realiza un mapeado o correspondencia de los campos de metadatos de los distintos esquemas que éstos emplean. Si además los repositorios comparten el mismo esquema se puede buscar por campos comunes.

Éste recolector resulta sencillo de instalar y configurar, empleando una base de datos MySQL o PostgreSQL, un servidor web Apache o Microsoft IIS, y pudiéndose instalar en cualquier sistema operativo incluido Windows. Asimismo, es uno de los sistemas más recientes y de los pocos que se actualiza todavía (su última versión, la 2.0.1, es de enero de 2007).

Todas estas razones contribuyen a que gran parte de los proveedores de servicios OAI de tamaño medio que se han estudiado en el capítulo 5 (véase 5.3. *Proveedores de servicios*) hayan sido contruidos con PKP Harvester, como: los temáticos Holmes y SDL (Search Digital Libraries) en biblioteconomía y documentación, o SEED del Instituto de Tecnología de la India, DOKs Harvester de los Institutos de Educación Superior de Flandes (Bélgica) o Repositorium (Venezuela) entre otros ejemplos.

8.4.4. Perl Harvester de Virginia Tech

Perl Harvester⁵⁹ es un recolector orientado a objetos que soporta las versiones 1.0, 1.1 y 2.0 del protocolo OAI-PMH, así como su versión extendida XOAI-PMH v1.0

⁵⁹ Información y paquetes de descarga de Perl Harvester en: <<http://oai.dlib.vt.edu/odl/software/harvest/>>. [Fecha de consulta: 10/07/2007].

(eXtended Protocol for Metadata Harvesting), requisito para la construcción de bibliotecas digitales abiertas (Open Digital Libraries, ODLs)⁶⁰. Este recolector en lenguaje Perl es un desarrollo del DLRL (Digital Library Research Laboratory) de Virginia Tech, y su última versión es la 2.0, de agosto de 2002.

El recolector de DLRL no requiere ningún tipo de proceso de instalación, tan sólo precisa de los módulos estándar de Perl y de un servidor capaz de funcionar con guiones CGI, y posteriormente es necesario copiar los guiones Perl en el directorio que corresponda. Una misma implementación puede ser utilizada fácilmente para recolectar metadatos de distintos sitios con propósitos diferentes. Todas las extensiones, configuraciones y contenedores se especifican empleando XML Schema.

8.4.5. Recolectores de la UIUC

La Universidad de Illinois en Urbana-Champaign cuenta con una herramienta para la recolección de metadatos o *harvesting*⁶¹, con dos versiones: una programada en Java (Java OAI Harvester), y otra en Visual Basic (VB OAI Harvester). Además, se ofrece una librería de objetos para este último (VB OAI Harvester Object Library). Estas herramientas pueden ser descargadas de Sourceforge junto al resto de las herramientas software desarrolladas por el grupo como soporte a la iniciativa OAI, y que se distribuyen bajo licencia de código abierto de NCSA / UIUC, con fecha de última actualización en 2004.

La versión del recolector en Visual Basic está diseñada para su instalación en una plataforma Windows 2000 o NT y precisa de un Microsoft Internet Information Server (IIS). Su arquitectura se compone de módulos ASP incluyendo VBScript, Microsoft XML Parser (MSXML) 4.0 Core Services y otros componentes Java, como Java Runtime Environment 1.3 o 1.4.

En cuanto al recolector programado en Java, es recomendable su instalación en un sistema Linux Red Hat 7.0 o superior, u otro sistema Linux compatible. Precisa de un servidor Apache v.1.3, Yakarta Tomcat 4.0 para los servlet Java, y otras herramientas Java (Java 2 SDK versión 1.4 o el Paquete Java XML).

Junto a estos recolectores se distribuye el paquete "VB OAI Harvester Object Library", que contiene una biblioteca de enlace dinámico (DLL) ActiveX creada para escribir otros programas de recolección de metadatos en repositorios OAI, compatible con todas sus versiones.

⁶⁰ Open Digital Libraries (ODLs) <<http://oai.dlib.vt.edu/odl/>>, un proyecto del DLIB de Virginia Tech. [Fecha de consulta: 10/07/2007]

⁶¹ Las herramientas de recolección de metadatos de la Universidad de Illinois en Urbana-Champaign se recogen en: <<http://nrgal.grainger.uiuc.edu/harvester.htm>>. [Fecha de consulta: 10/07/2007]

8.4.6. Net::OAI::Harvester

Net::OAI::Harvester es una extensión en Perl para consultar repositorios OAI-PMH de forma sencilla y recolectar metadatos de éstos. Desarrollado por los programadores Ed Summers y Martin Emmerich [Summers, 2004], se puede descargar su versión 1.0 (de julio de 2005) en el archivo de recursos en Perl CPAN (Comprehensive Perl Archive Network)⁶².

En cuanto a las características específicas de Net::OAI::Harvester, destacan su empleo de un analizador o parser (XML::SAX) para dividir los resultados de una consulta OAI-PMH de lista con posible respuesta larga, como ListRecords. Al dividir el documento XML de la respuesta se crean una serie de objetos Perl (registros, cabeceras...), que se almacenan en el disco duro de forma seriada y facilitan que se utilice una menor memoria de procesamiento. Además, los filtros XML::SAX permiten que cada implementador desarrolle sus propios paquetes para la indización de metadatos, de manera que acepten no sólo metadatos en Dublin Core sin cualificar sino cualquier otro formato.

8.4.7. OAIHarvester2

El proyecto OAIHarvester2 es una aplicación en Java que proporciona un recolector de metadatos de repositorios que cumplen OAI-PMH v.1.1 y v2.0. Es un desarrollo de software libre de la OCLC⁶³, que se distribuye de forma gratuita mediante licencia Apache versión 2.

En palabras de su desarrollador⁶⁴, Jeffrey Young, OAIHarvester2 resulta más ligera que su predecesora. Puede correr como una aplicación independiente o integrarse en aplicaciones existentes. Mientras que la antigua aplicación tenía un interfaz en Java que permitía el procesamiento de los registros recolectados sobre la marcha, la nueva versión simplemente agrupa las respuestas OAI en un archivo XML para su posterior procesamiento (p. ej. con XSLT).

⁶² Net::OAI::Harvester se puede descargar del repositorio CPAN: <<http://search.cpan.org/~esummers/OAI-Harvester-1.0/lib/Net/OAI/Harvester.pm>>. [Fecha de consulta: 10/07/2007].

⁶³ OAIHarvester2, de la OCLC: <<http://www.oclc.org/research/software/oai/harvester2.htm>>. [Fecha de consulta: 10/07/2007].

⁶⁴ Afirmación extraída de un correo electrónico de Jeffrey Young a la lista OAI-Implementers para anunciar la herramienta OAIHarvester2 en octubre de 2003: <<http://www.openarchives.org/pipermail/oai-implementers/2003-October/001081.html>>. [Fecha de consulta: 10/07/2007].

8.5. OTRAS HERRAMIENTAS QUE SOPORTAN EL PROTOCOLO OAI-PMH

En esta última sección se han recogido, a modo de ejemplo, un conjunto de herramientas muy distintas entre sí: una aplicación para la agregación y seguimiento de repositorios como Celestial, software de bibliotecas digitales que soportan el protocolo OAI; módulos que aportan funcionalidades OAI a distintos sistemas de gestión de contenido; plataformas para la publicación de revistas abiertas y conferencias; herramientas que permiten crear repositorios estáticos y pasarelas a los mismos; y por último, una herramienta que permite crear proveedores de datos a partir de bases de datos locales cualquiera mediante la transformación de su estructura de campos particular a los elementos Dublin Core sin cualificar.

El hecho de que distintos tipos de sistemas de gestión de contenido digital se vayan adaptando progresivamente para soportar el protocolo OAI-PMH nos da una idea de su enorme capacidad y potencialidad para el intercambio de datos entre múltiples y heterogéneos sistemas digitales de información, y en definitiva, para mejorar y simplificar la difusión y recuperación de la información electrónica a todos los niveles.

Por otro lado, y aunque no se hayan recogido aquí por su gran variedad y tipología, es necesario mencionar la existencia de una serie de herramientas, guiones u hojas de estilo que permiten transformar ficheros en un determinado formato de metadatos (MARC, TEI, LOM)⁶⁵ al estándar Dublin Core sin cualificar, es decir, el formato que determina el protocolo OAI-PMH y que se reconoce con las siglas oai_dc. De alguna manera y mediante el uso combinado con otras herramientas de proveedores de datos que se han descrito en este capítulo, estas herramientas también facilitan el cumplimiento de los objetivos del protocolo OAI-PMH en la difusión de contenidos digitales en Internet.

Y en especial, no debe obviarse una práctica que comienza a extenderse entre los productores de software de automatización de bibliotecas, y que consiste en el soporte al protocolo OAI-PMH en sus sistemas de gestión bibliotecaria. Ya lo han llevado a cabo algunas empresas internacionales como Sirsi Dynix con su producto Unicorn⁶⁶, Geac con Vubis Smart Library 2.4.1⁶⁷, Ex Libris con el sistema ALEPH 500⁶⁸, o VTLS con su

⁶⁵ Sirva de ejemplo MARCXML <<http://www.loc.gov/standards/marcxml/>> un conjunto de aplicaciones, hojas de estilo y otras utilidades, desarrolladas por la Biblioteca del Congreso de los Estados Unidos para el trabajo con MARC en entornos XML. [Fecha de consulta: 10/07/2007].

⁶⁶ En la información técnica de Unicorn se indica su cumplimiento con el protocolo OAI-PMH, entre otros estándares. *Sirsi Dynix Unicorn Library Management System. Technical Overview*. p. 6. Disponible en: <http://www.sirsidyndix.com/Resources/Pdfs/Solutions/Products/Unicorn_Technical_Overview.pdf> [Fecha de consulta: 10/07/2007].

⁶⁷ Como indica la nota de lanzamiento de la versión: *Geac Releases Version 2.4.1 of its Vubis Smart Library Automation System*, dic. 2005. Disponible en: <<http://www.vubis-smart.com/content.asp?PageID=149>>. [Fecha de consulta: 10/07/2007]

producto software libre VORTEX⁶⁹ que hace posible que los registros bibliotecarios estén disponibles a través del protocolo de recolección de metadatos OAI. En España destaca la empresa DIGIBIS, cuyo producto de automatización de bibliotecas DIGIBIB⁷⁰, cuenta con un módulo independiente de repositorio OAI así como un recolector OAI que se puede integrar con el sistema. Y en Venezuela, como ya se ha descrito, cuentan con el sistema de automatización de bibliotecas Alejandría (<http://alejandria.hacer.ula.ve>), para el que se ha desarrollado un módulo específico que convierte sus registros a OAI.

8.5.1. Seguimiento y monitorización de proveedores de datos: Celestial

Creado por Tim Brody en la Universidad de Southampton, Celestial (<http://celestial.eprints.org>) es una herramienta que permite la agregación, seguimiento y monitorización de repositorios que cumplan el protocolo OAI-PMH v.2.0. Antes denominado OAIA, es empleado por la iniciativa Eprints como base de su registro de archivos abiertos ROAR.

La estructura de Celestial se compone de la librería oai-perl v.2⁷¹, de una base de datos MySQL, Perl v.5.6, y de un servidor web como Apache, con capacidad de gestionar guiones CGI. Una vez instalado permite a su administrador seguir la evolución de los repositorios que se hayan seleccionado y agregado previamente. Se lanzan recolecciones de forma periódica en las que se recogen datos como el esquema de metadatos de los registros, fecha y hora de la última recolección, número total de registros en el archivo y tamaño del mismo en megabytes o kilobytes.

8.5.2. Software para bibliotecas digitales: Greenstone y otros

Al comienzo del capítulo se mencionaban diversas plataformas comerciales para la creación de bibliotecas digitales que soportan el protocolo OAI-PMH, en muchos casos mediante el desarrollo de módulos adicionales o add-ons, como CONTENTdm, Digital Commons, Digitool de Exlibris, Curator de Endeavor, Archivemaker de digitAlexandria,

⁶⁸ Véase al respecto la información sobre el producto ALEPH 500 <<http://www.exlibrisgroup.com/aleph.htm>> en el sitio web de la empresa Ex Libris. [Fecha de consulta: 10/07/2007].

⁶⁹ Véase al respecto la información sobre el producto VORTEX <<http://www.vtls.com/Products/vortex.shtml>> en el sitio web de la empresa de software bibliotecario VTLS (Visionary Technology Library Solutions). [Fecha de consulta: 10/07/2007].

⁷⁰ Información extraída de la página web de Digibis sobre el producto DIGIBIB, Sistema Integrado de Gestión Bibliotecaria: <<http://www.digibis.com/bibliotecas.htm>>. [Fecha de consulta: 10/07/2007].

⁷¹ Oai-perl library <<http://oai-perl.sourceforge.net/>> es una librería en lenguaje PERL compilada por la Universidad de Southampton que permite el desarrollo rápido de un interfaz compatible con OAI-PMH para un servidor o base de datos existente. [Fecha de consulta: 10/07/2007].

DLXS de la Universidad de Michigan o Intralibrary de Intrallect. Por supuesto, también existen algunos sistemas de bibliotecas digitales basados en software libre que, con el auge del movimiento de archivos abiertos, se han adaptado para cumplir con el protocolo OAI-PMH y otros estándares de interoperabilidad, por ejemplo: Greenstone, iVia, Keystone o CWIS.

Una de las herramientas más importantes y utilizadas a nivel internacional es Greenstone (<http://www.greenstone.org/>), un proyecto de la Universidad de Waitako (Nueva Zelanda). Greenstone permite tanto crear un proveedor de datos a partir de sus registros como recolectar registros de otras fuentes e incluirlos en su colección. Incluso, sus registros pueden ser exportados en METS y en el formato DSpace para ser directamente importados en un repositorio Dspace, de la misma manera que cualquier colección DSpace puede ser importada en Greenstone. En cuanto a los formatos de metadatos, aunque por defecto acepta metadatos Dublin Core (calificado y sin calificar), RFC 1807, NZGLS (New Zealand Government Locator Service) y AGLS (Australian Government Locator Service), su editor de metadatos (Greenstone's Metadata Set Editor) cuenta con varios *plug-ins* para adaptarse a los formatos XML, MARC, CDS/ISIS, ProCite, BibTex, Refer, OAI, DSpace y METS.

Otro ejemplo de software de biblioteca digital con funcionalidades OAI es el del sistema iVia Virtual Library Software (<http://ivia.ucr.edu/>), un desarrollo de software libre del Grupo de Investigación y Desarrollo iVia de la Universidad de California, Riverside, para el proyecto INFOMINE (<http://infomine.ucr.edu/>). iVia cuenta con un módulo propio, libiViaOaiPmh (una librería en C++), que permite implementar tanto proveedores de datos como de servicios que empleen el protocolo OAI-PMH.

Keystone Digital Library Suit (<http://www.indexdata.dk/keystone/>) es una familia de sistemas de código abierto para la gestión de contenido digital, gestión de portales y recuperación de información, que se distribuyen en un único paquete, y que permiten a las bibliotecas, museos y archivos crear servicios de biblioteca digital con funcionalidades OAI. Keystone permite establecer tanto un proveedor de datos como un proveedor de servicios. Como proveedor de datos acepta múltiples formatos de metadatos dependiendo de la existencia de hojas de estilo XSLT para la conversión de metadatos.

En cuanto al Collection Workflow Integration System (CWIS), es un paquete de software libre desarrollado por el Internet Scout Project (<http://scout.wisc.edu/>) en el seno de la iniciativa NSDL, que se adapta especialmente bien al trabajo con colecciones de menor tamaño. Incluye un motor de búsqueda, un sistema de evolución y recomendación de recursos, así como servidores OAI y RSS. En cuanto a los formatos de metadatos, soporta tanto el formato `oai_dc` como un perfil de aplicación propio adaptado de Dublin Core, `nsdl_dc`.

8.5.3. Sistemas de gestión de contenidos (CMS): OAI-PMH Pack

OAI-PMH Pack⁷² se trata de un conjunto de herramientas creadas por la empresa de software Infrac como extensiones a los sistemas Zope, Silva y Railroad y repositorios desarrollados en el lenguaje de programación código abierto Python, que les proporcionan facilidades basadas en el protocolo OAI-PMH. Sus componentes, todos ellos basados en Python, son: oaipmh Python module, OAICore Zope Product, SilvaOAI Silva Extension y OAI-PMH Railroad service, y pueden ser descargados y empleados de forma independiente⁷³.

SilvaOAI es una extensión al sistema de gestión de contenido de código abierto Silva (<http://www.infrac.com/products/silva>), que permite a los usuarios del sistema buscar y navegar por metadatos recolectados, de manera que una organización que emplee este gestor de contenido puede convertirse fácilmente en un proveedor de servicios OAI-PMH. Al tiempo, desarrolló un módulo para acceder a repositorios conformes a OAI-PMH programados en Python empleando un sencillo interfaz web. Este módulo soporta el conjunto de metadatos Dublin Core (*oai_dc*) y puede ser fácilmente extendido a otros formatos de metadatos. Otro componente del paquete es un sistema sofisticado de indexación y recolección de metadatos para emplear metadatos recolectados en el sistema de gestión de contenidos Zope (<http://www.zope.org>), el OAICore Zope Product. Ambos se han creado como bloques independientes que pueden aplicarse en otras aplicaciones basadas en Python o en Zope.

Por último, Infrac ha creado una extensión para Railroad (<http://www.infrac.com/products/railroad>) que le permite funcionar como un servidor OAI-PMH. Railroad es un repositorio de contenido basado en estándares para documentos digitales multimedia y sus metadatos correspondientes. Está diseñado para integrarse fácilmente con otros sistemas de gestión del contenido y otros software clientes, lo que ya se ha realizado con éxito con Silva y Plone (<http://www.plone.org>). Empleando Railroad y Silva de forma conjunta, una organización puede convertirse fácilmente en un proveedor de datos y en un proveedor de servicios OAI-PMH.

8.5.4. Sistemas de publicación de revistas y gestión de conferencias: OJS y OCS, de Public Knowledge Project

El Public Knowledge Project (<http://pkp.sfu.ca>), un proyecto de investigación conjunto de las universidades canadienses de British Columbia y Simon Fraser, ha diseñado diversas herramientas software libre que adoptan el protocolo OAI-PMH: una plataforma

⁷² OAIPack, de Infrac <<http://www.infrac.com/products/oaipack>>. Infrac es una compañía de desarrollo de software a petición de otras públicas o privadas generalmente europeas, especializada en gestión del contenido, autoría web, gestión de activos digitales y diseño de interfaces, y que implementa software de código abierto. [Fecha de consulta: 10/07/2007].

⁷³ No obstante, para instalar SilvaOAI es necesario haber instalado previamente OAICore Zope, el cual requiere a su vez de oaipmh Python.

web para la creación y difusión de revistas abiertas, OJS (Open Journal System); un sistema para la gestión de conferencias y eventos científicos y la publicación de actas, el OCS (Open Conferences System); y un recolector y proveedor de servicios, PKP Harvester.

El objetivo de estas herramientas es facilitar el proceso de distribución y publicación de información científica, mejorando la gestión y reduciendo los costes. Se pretende que estos sistemas creados por PKP, junto con una plataforma de creación de repositorios como DSpace o GNU Eprints, se empleen de forma integrada en una biblioteca universitaria y faciliten a los autores de la institución un servicio para el depósito de sus trabajos, y ofrezcan a toda la comunidad universitaria un punto de acceso único a dicha producción científica.

Tanto OCS como OJS han sido programadas en PHP, funcionan sobre servidores en sistemas operativos UNIX (OJS también en servidores Windows a partir de la versión 2.0.2), y precisan una base de datos MySQL o PostgreSQL. Ambas herramientas cuentan con numerosas implementaciones en todo el mundo: en marzo de 2007 OJS contaba con más de 900 títulos y OCS con unas setenta conferencias registradas.

8.5.5. Repositorios estáticos y pasarelas a repositorios estáticos

OAI Static Repository y OAI Static Repository Gateway⁷⁴ son dos aplicaciones de software libre desarrollada por LANL (Los Álamos National Laboratory) para proporcionar soporte al modelo de repositorios estáticos y pasarelas a repositorios estáticos, creados por la OAI para hacer aún más sencilla la exposición de los metadatos.

El “OAI Static Repository” de LANL es una solución sencilla para que los proveedores de datos pongan sus colecciones de metadatos a disposición del resto del mundo. Esta aplicación genera un fichero XML con un formato específico (un repositorio OAI estático) que hace que sea recolectable mediante el protocolo OAI-PMH a través de la intermediación de un sistema software (una pasarela a repositorios estáticos OAI) operado por un tercero.

La “OAI Static Repository Gateway” está programada en C, corre sobre cualquier plataforma Linux o Solaris, se distribuye bajo licencia BSD (Berkeley Software Distribution)⁷⁵ y está disponible para su descarga la versión 2.1.5. desde agosto de 2005.

La arquitectura de la OAI Static Repository Gateway de LANL, cuenta con cuatro componentes [Hochstenbach et al., 2003]:

- *Interfaz OAI-PMH*: un programa CGI que acepta consultas OAI-PMH dirigidas a repositorios estáticos, realiza un test de actualización (*freshness*) de los repositorios estáticos registrados, devuelve respuestas OAI-PMH en

⁷⁴ Información sobre el proyecto y software OAI Static Repository y OAI Static Repository Gateway en Sourceforge.net: <<http://srepor.sourceforge.net/>>. [Fecha de consulta: 10/07/2007].

⁷⁵ Licencia BSD (Berkeley Software Distribution): <<http://www.opensource.org/licenses/bsd-license.php>>. [Fecha de consulta: 10/07/2007].

caso de que un repositorio estático esté determinado para ser actualizado, genera los códigos de estado HTTP apropiados cuando falla el test de actualización y comunica la necesidad de actualización de la versión almacenada del repositorio al demonio (Daemon) a través de la Zona Segura (The Lock Zone).

- *Caché*: espacio de almacenamiento basado en un sistema de ficheros en el que las distintas versiones almacenadas (*cached*) de cada repositorio estático se gestionan en bases de datos GNU separadas.
- *Zona cerrada* (The Lock Zone): espacio de almacenamiento basado en un sistema de ficheros que actúa como una ventanilla entre el interfaz OAI-PMH y el demonio. Almacena ficheros cerrados, cada uno de ellos contiene información sobre un repositorio estático que necesita ser traído como resultado de un test de actualización fallido así como el estado actual del proceso de recogida.
- *Demonio* (Daemon): monitoriza de forma continuada la zona cerrada, va a buscar los repositorios estáticos cuando la zona cerrada lo requiere, actualizada el estado de recogida de los repositorios en los ficheros cerrados y actualiza la caché.

Para comprobar el funcionamiento de la OAI Static Repository Gateway, LANL ha implementado una pasarela experimental⁷⁶, con algunos repositorios estáticos de prueba ya registrados y con la opción de dar de alta el repositorio estático que se quiera.

8.5.6. Pasarela a bases de datos: FileMakerPro Repository Gateway

La pasarela a repositorios OAI de FileMakerPro⁷⁷, desarrollado en la Universidad de Illinois en Urbana-Champaign, es un servicio similar a la aplicación anterior, OAI Static Repository Gateway, que permite a las bases de datos FileMakerPro que ya estén accesibles en la web que puedan ser recolectadas mediante el protocolo OAI-PMH. Con este objetivo, se basa en la especificación de la Iniciativa de Archivos Abiertos sobre repositorios estáticos y pasarelas a repositorios estáticos.

Esta herramienta ha aprovechado la capacidad que FileMaker Pro ofrece desde su versión 5⁷⁸, de exportar versiones XML de las bases de datos en la web. De esta manera, OAI FileMakerPro Repository Gateway trata estos ficheros XML de manera muy similar a un fichero XML de un repositorio OAI estático. La única diferencia significativa es que el esquema XML de los registros que devuelve FileMaker puede variar

⁷⁶ "Experimental OAI Static Repository Gateway", Research Library of the Los Alamos National Laboratory: <<http://purl.lanl.gov/NET/srepod>>. [Fecha de consulta: 10/07/2007].

⁷⁷ OAI FileMakerPro Repository Gateway: <<http://ciccharvest.grainger.uiuc.edu/fmpgateway/>>. [Fecha de consulta: 10/07/2007].

⁷⁸ Más información sobre la capacidad XML de FileMakerPro en su página comercial: <http://www.filemaker.com/downloads/pdf/xml_overview.pdf> [Fecha de consulta: 10/07/2007].

dependiendo de la base de datos con la que se trabaje. Por lo tanto, se precisa una hoja de estilo XSLT para cada base de datos que permita transformar los registros XML de la base de datos FileMaker en un registro `oai_dc` como requiere el protocolo OAI-PMH.

8.5.7. Conversión de metadatos a `oai_dc`: Metadata Migrator

Desarrollada en PHP por la Universidad de Emory, con financiación del ILMS de Estados Unidos, Metadata Migrator (<http://www.MetaScholar.org/sw/mm>) es una herramienta libre de código abierto diseñada para facilitar el proceso de conseguir acceso vía OAI-PMH a registros en bases de datos locales creadas con herramientas como Procite o Filemaker.

Metadata Migrator permite transformar automáticamente los nombres de campos de los registros de las bases de datos de cada institución al formato Dublin Core sin cualificar, tal y como determina el protocolo OAI-PMH. Para este proceso se pide al usuario que indique uno a uno qué nombres de campos corresponden con cada elemento del esquema de metadatos Dublin Core.

REFERENCIAS BIBLIOGRÁFICAS

- CROW, Raym (2004). *A Guide to Institutional Repository Software* [en línea]. New York: Open Society Institute. 3ª ed. Disponible en: <http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf>. [Fecha de consulta: 10/07/2007].
- HALBERT, Martin y MILEWICZ, Elizabeth (2005). *OAI "Cheat Sheet" A taxonomy of Rapid OAI Deployment STRATEGIES* [en línea]. V. 11. Última revisión: 29 de noviembre de 2005. Disponible en: <<http://www.diglib.org/architectures/oai/impls2004/training/TaxonomyFinal.pdf>>. [Fecha de consulta: 25/06/2007]
- HOCHSTENBACH, Patrick; JEREZ, Henry; VAN DE SOMPEL, Herbert (2003). "The OAI-PMH Static Repository and Static Repository Gateway" [en línea]. En *ACM/IEEE-CS Joint Conference on Digital Libraries (Houston, Texas, 2003)*, p. 210-217. Disponible en: <<http://public.lanl.gov/herbertv/papers/jcdl2003-submitted-draft.pdf>> [Fecha de consulta: 10/07/2007].
- LAOMING, Xiu (2001). "DP9 in brief". *Dlib Magazine* [en línea], vol. 7, n. 12. Disponible en: <<http://www.dlib.org/dlib/december01/12inbrief.html>>. [Fecha de consulta: 10/07/2007]
- SILIÓ, Teresa (2005). "Fundamentos tecnológicos del acceso abierto: Open Archives Initiative y Open Archival Information System". *El profesional de la información* [en línea], vol. 14, n. 5, p. 365-380. Disponible en: <<http://eprints.rclis.org/archive/00005887/01/EPITSilio.pdf>>. [Fecha de consulta: 10/07/2007].
- SUMMERS, Ed (2004). "Building OAI-PMH Harvesters with Net::OAI::Harvester". *Ariadne* [en línea], n. 38. Disponible en: <<http://www.ariadne.ac.uk/issue38/summers>>. [Fecha de consulta: 30/02/2007].
- VILLEGAS ORTEGA, Isaac Abraham (2005). *Voai: un generador automático de servidores de metadatos bajo el protocolo OAI-PMH*. Tesis de Licenciatura. Puebla (México): Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de Las Américas, en Puebla. p. 19-24. Disponible en: <http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/villegas_o_ia/> [Fecha de consulta: 10/07/2007].